

STIC-Biotech/ChemLib

170169

mg

From: Marvich, Maria
Sent: Tuesday, November 01, 2005 6:33 AM
To: STIC-Biotech/ChemLib
Subject: 09/463542

Please search including interference 1-125 of SEQ ID NO:1
818-1320 of SEQ ID NO:3

368--1144 of SEQ ID NO:34

Thank you

Maria Bonovich Marvich, PhD
United States Patent and Trademark Office
Remsen 2B84
AU 1636
Mail Box 2C70
571-272-0774

1633?

RECEIVED
NOV - 1 2005
STIC/CHEM DIVISION

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: 11/1/05
Date completed: 11/1/05
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# 3 AA#: _____
S/L: 2 Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: 04
WWW/Internet: _____
Other (Specify): _____

THIS PAGE BLANK (USPTO)


```
Query Match      94.0%; Score 473; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 5e-101;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CAGTAGCATGCTGATACCAAGCTTTAAACTATGAGATACATATTTGAATTCAAAATTTTTC 90
    |||||
Db 1 CAGTAGCATGCTGATACCAAGCTTTAAACTATGAGATACATATTTGAATTCAAAATTTTTC 60

QY 91 TTCAGATATCTGATTAGAGATTAGAGATTCAACCCAGGGATAGACACCGAAGAAACTT 150
    |||||
Db 61 TTCAGATATCTGATTAGAGATTAGAGATTCAACCCAGGGATAGACACCGAAGAAACTT 120

QY 151 TGCCCAATAAGCTTTCTGGTATTTTCAATGCAAGAGATTTAAAGTTTCCATTTAAGAG 210
    |||||
Db 121 TGCCCAATAAGCTTTCTGGTATTTTCAATGCAAGAGATTTAAAGTTTCCATTTAAGAG 180

QY 211 CCATTGTGAATTATACAAACAATAAATAAGCAAGTGAGATTTGAACAGTCTCTCTCTGA 270
    |||||
Db 181 CCATTGTGAATTATACAAACAATAAATAAGCAAGTGAGATTTGAACAGTCTCTCTCTGA 240

QY 271 TAATTCTAAATACAGTACAGTTCACGCCCTCAGGAGACACTGAACTGTGGTCACCGGC 330
    |||||
Db 241 TAATTCTAAATACAGTACAGTTCACGCCCTCAGGAGACACTGAACTGTGGTCACCGGC 300

QY 331 GAGACAGTGTGGCAATATTATCCCTGTAATGTACCAAGTCTTGGCAGAGCAGTGAACATT 390
    |||||
Db 301 GAGACAGTGTGGCAATATTATCCCTGTAATGTACCAAGTCTTGGCAGAGCAGTGAACATT 360

QY 391 ATGACACAACCTTTTGTGTCAGAGTGGCTCCCTTAATAGGACAGTGCAGGCAATTTCAAGCCC 450
    |||||
Db 361 ATGACACAACCTTTTGTGTCAGAGTGGCTCCCTTAATAGGACAGTGCAGGCAATTTCAAGCCC 420

QY 451 AGTCCCTTCTGTGTTTATCCCATCTCTCCCAATATTTTGGAACTGATGCT 503
    |||||
Db 421 AGTCCCTTCTGTGTTTATCCCATCTCTCCCAATATTTTGGAACTGATGCT 473

RESULT 2
AF310249 984 bp DNA linear PRI 12-MAR-2001
LOCUS Homo sapiens peroxisome proliferator activated receptor gamma 2
DEFINITION gene, upstream sequence.
ACCESSION AF310249
VERSION AF310249.1 GI:13274397
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Lim,C.H., Kok,C.C., Samian,M.R., Najimudin,N. and Tengku
Muhammad,T.S.
JOURNAL Molecular cloning and sequencing of the human Peroxisome
Proliferator Activated Receptor Gamma 2 promoter
REFERENCE Unpublished
AUTHORS Lim,C.H., Kok,C.C., Samian,M.R., Najimudin,N. and Tengku
Muhammad,T.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2000) School of Biological Sciences, Universiti
Sains Malaysia, Minden, Penang 11800, Malaysia
FEATURES
source 1. .984
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
misc_feature 1. .982
/note="contains promoter and 5' untranslated region of
peroxisome proliferator activated receptor gamma 2 gene"
<982..>984
/product="peroxisome proliferator activated receptor gamma
2"
CDS 982..>984
```

```
/codon_start=1
/product="peroxisome proliferator activated receptor gamma
2"
/protein_id="AAK17925.1"
/db_xref="GI:13274398"
/translation="M"

ORIGIN
Query Match      90.5%; Score 455.2; DB 9; Length 984;
Best Local Similarity 96.8%; Pred. No. 7.4e-97;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

QY 1 TAGAGTAGTAGTACCTTAGGAATATACCAATTTTCAGTAGCATGCTGATACCAAGCTTTAAACT 60
    |||||
Db 362 TAGAGTAGTAGTACCTTAGGAATATACCAATTTTCAGTAGCATGCTGATACCAAGCTTTAAACT 421

QY 61 ATGAGTACATATTTGAAATTTCCAAATTTTCTTCAGATAAATGTGTGATTTAGAGATTTAGAGATT 120
    |||||
Db 422 ATGAGTACATATTTGAAATTTCCAAATTTTCTTCAATAATGT-----GATTAGAGATT 474

QY 121 CAACCCAGGGATAGACACCGGAAGAAACTTTGCCCAATTAAGCTTTCTGGTATTTTCATAA 180
    |||||
Db 475 CAACCCAGGGATAGACACCGGAAGAAACTTTGCCCAATTAAGCTTTCTGGTATTTTCATAA 534

QY 181 GCAAGAGATTTAAGTTTTCCATTTTAAGAAGCAATTTTGGCCCAATTAAGCTTTCTGGTATTTTCATAA 240
    |||||
Db 535 GCAAGAGATTTAAGTTTTCCATTTTAAGAAGCAATTTTGGCCCAATTAAGCTTTCTGGTATTTTCATAA 594

QY 241 CAAGTGTGATTTGAACAGTCTCTCTCTGATATAATCTTAAATACAGTACAGTTCACGCCCC 300
    |||||
Db 595 CAAGTGTGATTTGAACAGTCTCTCTCTGATATAATCTTAAATACAGTACAGTTCACGCCCC 654

QY 301 TCACGACACACTGACACATGT-GGTCACGGCGGACAGTGTGGCAATATTATCCCTGTAA 359
    |||||
Db 655 TCACGACACACTGACACATGTGGGT-CACGGCGGACAGTGTGGCAATATTATTCCTGTAA 714

QY 360 TGTACCAAGTCTTGCACAGAGCAGTGAACATTTATGACACAACTTTTGTTCACAGCTGGCTC 419
    |||||
Db 715 TGTACCAAGTCTTGCACAGAGCAGTGAACATTTATGACACAACTTTTGTTCACAGCTGGCTC 774

QY 420 CTAATAGGACAGTGCAGGCAATTTCAAGCCAGTCCCTTTCTGTGTTTATTCCTCATCTCTC 479
    |||||
Db 775 CTAATAGGACAGTGCAGGCAATTTCAAGCCAGTCCCTTTCTGTGTTTATTCCTCATCTCTC 834

QY 480 CCAAATATTTGGAACTGATGCTCT 503
    |||||
Db 835 CCAAATATTTGGAACTGATGCTCT 858

RESULT 3
AR121474 1100 bp DNA linear PAT 16-MAY-2001
LOCUS Sequence 10 from patent US 6159734.
DEFINITION AR121474
ACCESSION AR121474
VERSION AR121474.1 GI:14105050
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1100)
McKay,R., Borchers,A.H. and Baker,B.F.
TITLE Antisense modulation of peroxisome proliferator-activated receptor
gamma expression
JOURNAL Patent: US 6159734-A 10 12-DEC-2000;
FEATURES Location/Qualifiers
source 1. .1100
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match      90.5%; Score 455.2; DB 6; Length 1100;
Best Local Similarity 96.8%; Pred. No. 7.3e-97;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;
```

```
Qy 1 TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGTTTAAACT 60
Db 239 TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGTTTAAACT 298
Qy 61 ATGATACATATTTGAATTTCCAAATTTTCTTCAGATTAATGATTAGAGATTAGAGATT 120
Db 299 ATGATACATATTTGAATTTCCAAATTTTCTTCAGATTAATGATTAGAGATTAGAGATT 351
Qy 121 CAACAGGATAGACACCGAAAGAAACTTTGCCCAATAAAGCTTTCTGGTATTTCATAA 180
Db 352 CAACAGGATAGACACCGAAAGAAACTTTGCCCAATAAAGCTTTCTGGTATTTCATAA 411
Qy 181 GCAAGAGATTTAAGTTTCCATTTTAAGAGCAATTTGGAATTTTACAAATAAATAATG 240
Db 412 GCAAGAGATTTAAGTTTCCATTTTAAGAGCAATTTGGAATTTTACAAATAAATAATG 471
Qy 241 CAAGTGGATATTGAACAGTCTCTCTCTGATTAATTTCTAATAACAGTACAGTTCACGCC 300
Db 472 CAAGTGGATATTGAACAGTCTCTCTCTGATTAATTTCTAATAACAGTACAGTTCACGCC 531
Qy 301 TCACGAGACACTGAACATGT--GGTCACCGCGGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db 532 TCACGAGACACTGAACATGTGGTTCACCGCGGAGACAGTGTGGCAATATTATCCCTGTAA 591
Qy 360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTTATGACACAACTTTTGTGTCAGCTGGCTC 419
Db 592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTTATGACACAACTTTTGTGTCAGCTGGCTC 651
Qy 420 CTAAATAGACAGTGCACGCAATTCAGCCAGTCTTCTGTGTTATTTCCCATCTCTC 479
Db 652 CTAAATAGACAGTGCACGCAATTCAGCCAGTCTTCTGTGTTATTTCCCATCTCTC 711
Qy 480 CCAATATTTGGAACTGATGCT 503
Db 712 CCAATATTTGGAACTGATGCT 735
```

```
RESULT 4
LOCUS AB005520 1100 bp DNA linear PRI 10-NOV-1997
DEFINITION Homo sapiens ppar gamma2 gene for peroxisome proliferator
activated-receptor gamma, partial cds and 5' flanking.
```

```
AB005520
ACCESSION AB005520.1 GI:2605488
```

```
VERSION ppar gamma2; peroxisome proliferator activated-receptor gamma.
```

```
KEYWORDS Homo sapiens (human)
```

```
SOURCE Homo sapiens
```

```
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE 1 (sites)
```

```
AUTHORS Okazawa,H., Mori,H., Tamori,Y., Araki,S., Niki,T., Masugi,J.,
```

```
TITILE Kawanishi,M., Kubota,T., Sinoda,H. and Kasuga,M.
```

```
JOURNAL No coding mutations are detected in the peroxisome
```

```
REFERENCE lipotrophic diabetes
```

```
AUTHORS Diabestes (1997) In press
```

```
TITLE 2 (bases 1 to 1100)
```

```
JOURNAL Okazawa,H.
```

```
REFERENCE Direct Submission
```

```
AUTHORS Submitted (03-JUL-1997) Hideki Okazawa, Kobe University School of
```

```
JOURNAL Medicine, 2nd Department of Internal Medicine; 7-5-1 Kusunoki-cho
```

```
chuo-ku, Kobe 650, Japan [E-mail:okazawa@med.kobe-u.ac.jp,
```

```
FEATURES Tel:81-78-341-7451, Fax:81-78-382-2080)
```

```
source Location/Qualifiers
```

```
1..1100
```

```
/organism="Homo sapiens"
```

```
/mol_type="genomic DNA"
```

```
/db_xref="taxon:9606"
```

```
/tissue type="placenta"
```

```
859..940
```

```
gene /gene="ppar gamma2"
```

```
CDS 859..940
```

```
/gene="ppar gamma2"
/note="ppar gamma2 specific coding region"
/codon_start=1
/product="peroxisome proliferator activated-receptor
gamma"
/protein_id="BAA23353.1"
/db_xref="GI:2605488"
/translation="MGETLGDSPIDPDSFTDLSANISQ"
<859..940
/gene="ppar gamma2"
```

exon

ORIGIN

```
Query Match 90.5%; Score 455.2; DB 9; Length 1100;
Best Local Similarity 96.8%; Pred. No. 7.3e-97;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;
Qy 1 TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGTTTAAACT 60
Db 239 TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGTTTAAACT 298
Qy 61 ATGATACATATTTGAATTTCCAAATTTTCTTCAGATTAATGATTAGAGATTAGAGATT 120
Db 299 ATGATACATATTTGAATTTCCAAATTTTCTTCAGATTAATGATTAGAGATTAGAGATT 351
Qy 121 CAACAGGATAGACACCGAAAGAAACTTTGCCCAATAAAGCTTTCTGGTATTTCATAA 180
Db 352 CAACAGGATAGACACCGAAAGAAACTTTGCCCAATAAAGCTTTCTGGTATTTCATAA 411
Qy 181 GCAAGAGATTTAAGTTTCCATTTTAAGAGCAATTTGGAATTTTACAAATAAATAATG 240
Db 412 GCAAGAGATTTAAGTTTCCATTTTAAGAGCAATTTGGAATTTTACAAATAAATAATG 471
Qy 241 CAAGTGGATATTGAACAGTCTCTCTCTGATTAATTTCTAATAACAGTTCACGCC 300
Db 472 CAAGTGGATATTGAACAGTCTCTCTCTGATTAATTTCTAATAACAGTTCACGCC 531
Qy 301 TCACGAGACACTGAACATGT--GGTCACCGCGGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db 532 TCACGAGACACTGAACATGTGGTTCACCGCGGAGACAGTGTGGCAATATTATCCCTGTAA 591
Qy 360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTTATGACACAACTTTTGTGTCAGCTGGCTC 419
Db 592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTTATGACACAACTTTTGTGTCAGCTGGCTC 651
Qy 420 CTAAATAGACAGTGCACGCAATTCAGCCAGTCTTCTGTGTTATTTCCCATCTCTC 479
Db 652 CTAAATAGACAGTGCACGCAATTCAGCCAGTCTTCTGTGTTATTTCCCATCTCTC 711
Qy 480 CCAATATTTGGAACTGATGCT 503
Db 712 CCAATATTTGGAACTGATGCT 735
```

RESULT 5

LOCUS

AX951686/c

Sequence 3 from Patent WO0309310.

DEFINITION AX951686

AX951686.1 GI:40782040

VERSION Homo sapiens (human)

KEYWORDS Homo sapiens

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Bullerdiel,J.

AUTHORS Novel nucleic acid sequences and proteins of tumors and neoplasias

TITLE of the thyroid gland

JOURNAL Patent: WO 0309310-A 3 13-NOV-2003;

Universitaet Bremen (DE)

FEATURES Location/Qualifiers

source 1..85873

/organism="Homo sapiens"

DNA linear PAT 09-JAN-2004

ORIGIN	/mol_type="unassigned DNA"	
	/db_xref="taxon:9606"	
Query Match	90.5%; Score 455.2; DB 6; Length 85873;	
	Best Local Similarity 96.8%; Pred. No. 4.7e-97;	
Matches	488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;	
Qy	1	TAGAGTAGTACCTTAGGAATATACATTTTCAGTAGCAGTGTGATACCAAGTTTAACT 60
Db	54472	TAGAGTAGTACCTTAGGAATATACATTTTCAGTAGCAGTGTGATACCAAGTTTAACT 54413
Qy	61	ATGGATACATATTTGAATTCCTCAAAATTTTCTTCAGATAATGATGATAGAGATTAGAGATT 120
Db	54412	ATGGATACATATTTGAATTCCTCAAAATTTTCTTCAGATAATGATGATAGAGATT 54360
Qy	121	CAACAGGATAGACACCGAAGAAACTTTGGCCAAATAAGCTTTCTGGTATTTCATAA 180
Db	54359	CAACAGGATAGACACCGAAGAAACTTTGGCCAAATAAGCTTTCTGGTATTTCATAA 54300
Qy	181	GCAAGAGATTTAAGTTTCCATTTTAAGAGCCATTTGTAATATACACATTAATAAATG 240
Db	54299	GCAAGAGATTTAAGTTTCCATTTTAAGAGCCATTTGTAATATACACATTAATAAATG 54240
Qy	241	CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTAAATACAGTACAGTTACGCCCC 300
Db	54239	CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTAAATACAGTACAGTTACGCCCC 54180
Qy	301	TCACGAGACATGAACATGTT-GGTCCCGGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db	54179	TCACAAGACATGAACATGTTGGGTCCCGGAGACAGTGTGGCAATATTATTCCTCTGAA 54120
Qy	360	TGTACCAAGTCTGCCAGAGCAGTGAACATTTATGACACAACTTTTTCAGCTGGCTC 419
Db	54119	TGTACCAAGTCTTGCCTAAAGCAGTGAACATTTATGACACAACTTTTTCAGCTGGCTC 54060
Qy	420	CTAATAGACAGTGCAGCCAAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 479
Db	54059	CTAATAGACAGTGCAGCCAAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 54000
Qy	480	CCAAATATTGGAACTGATGCT 503
Db	53999	CCAAATATTGGAACTGATGCT 53976
RESULT 6	AC093174	135675 bp DNA linear PRI 08-NOV-2002
LOCUS	Homo sapiens chromosome 3 clone RP11-167M22 map 3p, complete sequence.	
ACCESSION	AC093174	
VERSION	AC093174.2	GI:24796717
KEYWORDS	HTG.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 135675)	
AUTHORS	Wu, Q., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, O., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	
	Chromosome 3p genomic sequence	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 135675)	
REFERENCE	Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,	
AUTHORS		
TITLE	Direct Submission	
	Submitted (08-NOV-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China	
JOURNAL	On Nov 8, 2002 this sequence version replaced gi:15148930.	
	-----Genome Center	
COMMENT	Center:Beijing Center	
	Center code:Beijing	
	Website:http://hg.cigtp.ac.cn	
	http://www.genomics.org.cn	
	Contact:hgc@cigtp.ac.cn	
	----- Project Information	
	Center project name:1k project	
	Center clone name: RP11-167M22	
	----- Summary Statistics	
	Sequencing vector: pUC18; 100% of reads	
	Chemistry: Dye-terminator: ET 55% of reads	
	Chemistry: Dye-terminator: Big Dye; 45% of reads	
	Assembly program: Phrap; version 0.990329	
	Consensus quality: 0 bases at least Q40	
	Consensus quality: 0 bases at least Q30	
	Consensus quality: 6 bases at least Q20	
	Insert size: 3392; sum-of-contigs	
	Quality coverage: 0.00x in Q20 bases,sum-of-contigs	
FEATURES	-----	
	Location/Qualifiers	
source	i. 135675	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="3"	
	/map="3p"	
	/clones="RP11-167M22"	
ORIGIN		
Query Match	90.5%; Score 455.2; DB 9; Length 135675;	
	Best Local Similarity 96.8%; Pred. No. 4.5e-97;	
Matches	488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;	
Qy	1	TAGAGTAGTACCTTAGGAATATACATTTTCAGTAGCAGTGTGATACCAAGTTTAACT 60
Db	81198	TAGAGTAGTACCTTAGGAATATACATTTTCAGTAGCAGTGTGATACCAAGTTTAACT 81257
Qy	61	ATGGATACATATTTGAATTCCTCAAAATTTTCTTCAGATAATGATGATAGAGATTAGAGATT 120
Db	81258	ATGGATACATATTTGAATTCCTCAAAATTTTCTTCAGATAATGATGATAGAGATT 81310

Qy	121	CAACCGGATAGACACCGAAGAACTTTGGCCAAATAAGCTTTCTGGTATTTCTATTA	180
Db	81311	CAACCGGATAGACACCGAAGAACTTTGGCCAAATAAGCTTTCTGGTATTTCTATTA	81370
Qy	181	GCAAGAGATTAAAGTTTTTCCATTTAAGAGCCATTGTGAATTATACAAATAAATAAATG	240
Db	81371	GCAAGAGATTAAAGTTTTTCCATTTAAGAGCCATTGTGAATTATACAAATAAATAAATG	81430
Qy	241	CAAGTGATATTGAACAGTCTCTTCTCTGATAATTTCTAAATACAGTACAGTTTCAACGCC	300
Db	81431	CAAGTGATATTGAACAGTCTCTTCTCTGATAATTTCTAAATACAGTACAGTTTCAACGCC	81490
Qy	301	TCACGACACTGAAACATGT-GGTCCACCGGAGACAGTGTGGCAATATATCCCTGTAA	359
Db	81491	TCACGACACTGAAACATGTGTGGTTCACCGGAGACAGTGTGGCAATATATCCCTGTAA	81550
Qy	360	TGTACCAAGTCTGCCAGACAGTGAACATTATGACACACTTTTGTACAGCTGGCTC	419
Db	81551	TGTACCAAGTCTGCCAGACAGTGAACATTATGACACACTTTTGTACAGCTGGCTC	81610
Qy	420	CTAATAGCAGTGCCAGCCAAATTCAGCCAGTCTCTTCTGTGTTTATTCCTCATCTCTC	479
Db	81611	CTAATAGCAGTGCCAGCCAAATTCAGCCAGTCTCTTCTGTGTTTATTCCTCATCTCTC	81670
Qy	480	CCAAATTTTGGAACTGATGCT	503
Db	81671	CCAAATTTTGGAACTGATGCT	81694
RESULT 7			
LOCUS	AY157024	148828 bp DNA linear	PRI 16-JUN-2004
DEFINITION	Homo sapiens peroxisome proliferative activated receptor gamma (PPARG) gene, complete cds.		
ACCESSION	AY157024	GI:48762804	
VERSION	AY157024.2	GI:48762804	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. (bases 1 to 148828)		
AUTHORS	Rieder, M.J., da Ponte, S.H., Kuldane, S.A., Rajkumar, N., Smith, J.D., Toth, E.J. and Nickerson, D.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-SEP-2002) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA		
REFERENCE	2. (bases 1 to 148828)		
AUTHORS	Rieder, M.J., Daniels, R.L., da Ponte, S.H., Hastings, N.C., Ahearn, M.O., Rajkumar, N., Yi, Q. and Nickerson, D.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-JUN-2004) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA		
REMARK	Sequence update by submitter		
COMMENT	On Jun 16, 2004 this sequence version replaced gi:23953882. To cite this work please use: SeattleSNPs. NHLBI Program for Genomic Applications, UW-FHCRC, Seattle, WA (URL: http://pga.mbt.washington.edu).		
FEATURES	Location/Qualifiers		
source	1. 148828		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
repeat_region	9..59		
	/rpt_family="L1"		
variation	54		
	/rpt_type="dispersed"		
	/frequency="0.12"		
repeat_region	60..351		
	/rpt_family="Alu"		
variation	147		
	/rpt_type="dispersed"		
	/frequency="0.01"		
	/replace="n"		
	/frequency="0.03"		
	/replace="t"		
variation	252		
	/frequency="0.03"		
	/replace="g"		
repeat_region	352..431		
	/rpt_family="L1"		
	/rpt_type="dispersed"		
variation	374		
	/frequency="0.05"		
	/replace="c"		
repeat_region	435..759		
	/rpt_family="Alu"		
	/rpt_type="dispersed"		
variation	1555		
	/frequency="0.02"		
	/replace="t"		
misc_feature	1754..1925		
	/note="Region not scanned for variation"		
repeat_region	1784..2127		
	/rpt_family="L1"		
	/rpt_type="dispersed"		
repeat_region	2032..2632		
	/rpt_family="L1"		
	/rpt_type="dispersed"		
gene	2048..148443		
	/gene="PPARG"		
mRNA			

```
variation
4160 /gene="PPARG"
    /frequency="0.01"
    /replace="c"
4220 /gene="PPARG"
    /note="Region not scanned for variation"
repeat_region
5952. .6004 /rpt_family="MIR"
    /rpt_type=dispersed
repeat_region
8735. .9019 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
9637. .9935 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
10092. .10404 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
10423. .10537 /rpt_family="MER1 type"
    /rpt_type=dispersed
repeat_region
11193. .11356 /rpt_family="L1"
    /rpt_type=dispersed
repeat_region
12140. .12449 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
12949. .13255 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
13277. .13383 /rpt_family="L2"
    /rpt_type=dispersed
repeat_region
16397. .16653 /rpt_family="L1"
    /rpt_type=dispersed
repeat_region
17063. .17196 /rpt_family="MER1 type"
    /rpt_type=dispersed
repeat_region
17325. .17622 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
17653. .17733 /rpt_family="L2"
    /rpt_type=dispersed
repeat_region
17766. .17984 /rpt_family="CR1"
    /rpt_type=dispersed
repeat_region
18787. .18845 /rpt_family="L2"
    /rpt_type=dispersed
repeat_region
19727. .19804 /rpt_family="L2"
    /rpt_type=dispersed
repeat_region
19757. .19813 /rpt_family="MIR"
    /rpt_type=dispersed
repeat_region
20083. .20181 /rpt_family="MIR"
    /rpt_type=dispersed
repeat_region
20182. .20524 /rpt_family="MER1 type"
    /rpt_type=dispersed
repeat_region
20525. .20660 /rpt_family="MIR"
    /rpt_type=dispersed
repeat_region
20937. .21239 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
21961. .22204 /rpt_family="L2"
    /rpt_type=dispersed
```

```
repeat_region
22205. .22492 /rpt_family="Alu"
    /rpt_type=dispersed
repeat_region
22493. .22581 /rpt_family="L2"
    /rpt_type=dispersed
repeat_region
22582. .22916 /rpt_family="MER1 type"
    /rpt_type=dispersed
repeat_region
22917. .23104 /rpt_family="L2"
    /rpt_type=dispersed

Query Match      90.5%; Score 455.2; DB 9; Length 148828;
Best Local Similarity 96.8%; Pred. No. 4.4e-97;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
Db TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 65152

Qy 61 ATGGATACATATTGGAATTTCCAAATTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
Db ATGGATACATATTGGAATTTCCAAATTTTCTTCAAATAATGT-----GATTAGAGATT 65205

Qy 121 CAACGAGGATAGACACCGAAAGAAACTTTGCCCAATAAGCTTCTCGGTATTTCATAA 180
Db CAACGAGGATAGACACCGAAAGAAACTTTGCCCAATAAGCTTCTCGGTATTTCATAA 65265

Qy 181 GCAAGAGATTAAAGTTTCCATTTTGAAGCCATTGTGAATTATACAAATAAAAAATG 240
Db GCAAGAGATTAAAGTTTCCATTTTGAAGCAATTGTGAATTTTACAAATAAAAAATG 65325

Qy 241 CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTAAATACAGTACAGTTCACGCC 300
Db CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTAAATACAGTACAGTTCACGCC 65385

Qy 301 TCACGAGACACTGAAACATGT-GGTCACCGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db TCACGAGACACTGAAACATGTGGGTACCGCGGAGACAGTGTGGCAATATTATCCCTGTAA 65445

Qy 360 TGTAACCAAGTCTTGCCAGAGCAGTGAACATATGACACAACTTTTTGTGTCAGTGGT 419
Db TGTAACCAAGTCTTGCCAAAGCAGTGAACATATGACACAACTTTTTGTGTCAGTGGT 65505

Qy 420 CTAATAGGACAGTCCAGCCCAATTTCAGCCAGTCCCTTTCTGTGTTTATCCCATCTCTC 479
Db CTAATAGGACAGTCCAGCCCAATTTCAGCCAGTCCCTTTCTGTGTTTATCCCATCTCTC 65565

Qy 480 CCAAAATTTTGGAAACTGATGTCT 503
Db CCAAAATTTTGGAAACTGATGTCT 65589
```

RESULT 8

AC090947/c

LOCUS

DEFINITION Homo sapiens chromosome 3 clone RP11-30G23 map 3p, complete

AC090947

AC090947 AC016333

VERSION AC090947.2 GI:24796728

KEYWORDS HTG; HTGS DRAFT.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 166049)
Wu, Q., Bao, J., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,
Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Guo, D.,
Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,
Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,
Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Luo, X., Qi, X., Song, L.,
Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,

Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,
Wang, Y., Wu, D., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,
Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, H., Zhang, M.,
Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
Yu, J. and Yang, H.

Chromosome 3p genomic sequence

Unpublished

2 (bases 1 to 166049)

Wu, Q., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,
Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,
Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,
Li, F., Li, G., Li, J., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,
Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L.,
Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,
Wang, J., Wang, L., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,
Wang, Y., Wu, D., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,
Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,
Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
Yu, J. and Yang, H.

Direct Submission

Submitted (20-MAR-2001) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China

3 (bases 1 to 166049)

Wu, Q., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,
Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,
Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,
Li, F., Li, G., Li, J., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,
Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L.,
Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,
Wang, J., Wang, L., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,
Wu, D., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y.,
Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X.,
Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and
Yang, H.

Direct Submission

Submitted (08-NOV-2002) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China

On Nov 8, 2002 this sequence version replaced gi:13384351.

-----Genome Center

Center:Beijing Center

Center code:Beijing

Website:http://hgsc.igtp.ac.cn

http://www.genomics.org.cn

Contact:hgsc@igtp.ac.cn

----- Project Information

Center project name:1k project

Center clone name:RP11-30G23

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator: ET 55% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 499 bases at least Q40

Consensus quality: 605 bases at least Q30

Consensus quality: 674 bases at least Q20

Insert size: 692; sum-of-contigs

Quality coverage: 2.80x in Q20 bases;sum-of-contigs

FEATURES

source

Location/Qualifiers
1. 166049
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3p"
/map="3p"
/clone="RP11-30G23"

ORIGIN

Query Match 90.5%; Score 455.2; DB 9; Length 166049;
Best Local Similarity 96.8%; Pred. No. 4.4e-97;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAACTACCTTAGGAATATATACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
145789 TAGAGTAACTACCTTAGGAATATATACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 145730
Db
Qy 61 ATGGATACATATTTGAATTCCTTCCAAATTTTCTTCAGATATGTTAGTAGAGATTAGAGATT 120
145729 ATGGATACATATTTGAATTTCCAAATTTTCTTCCAAATTAATGT-----GATTAGAGATT 145677
Db
Qy 121 CAACAGGAGATAGACACCGGAAAGAAACTTTGGCCCAATAAGCTTTCTGGTATTTCTATAA 180
145676 CAACAGGAGATAGACACCGGAAAGAAACTTTGGCCCAATAAGCTTTCTGGTATTTCTATAA 145617
Db
Qy 181 GCAAGAGATTAAAGTTTTCATTTTGAAGAGCCATTGGAATATATACAACTATAAATAATG 240
145616 GCAAGAGATTAAAGTTTTCATTTTGAAGAGCAATTTGGAATTTTACAACTATAAATAATG 145557
Db
Qy 241 CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTTAATACAGTACAGTTCACGCCCC 300
145556 CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTTAATACAGTACAGTTCACGCCCC 145497
Db
Qy 301 TCACGAGACACTGAACATGT-GGTCAACGCGGAGACAGTGTGGCAATATATCCCTGTAA 359
145496 TCACGAGACACTGAACATGTGGTTCACGCGGAGACAGTGTGGCAATATTTCCCTGTAA 145437
Db
Qy 360 TGTACCAAGTCTTCCAGAGAGAGCATGTAATATGACACAACTTTTGTCAAGCTGGCTC 419
145436 TGTACCAAGTCTTCCAGAGAGAGCATGTAATATGACACAACTTTTGTCAAGCTGGCTC 145377
Db
Qy 420 CTATAGACAGTGTCCAGCCCAATTTCAAGCCAGTCTCTCTGTTTATTTCCCATCTCTC 479
145376 CTATAGACAGTGTCCAGCCCAATTTCAAGCCAGTCTCTCTGTTTATTTCCCATCTCTC 145317
Db
Qy 480 CCAATATTTTGGAACTGATGTCT 503
145316 CCAATATTTTGGAACTGATGTCT 145293
Db
RESULT 9
AC151849/c 186028 bp DNA linear HTG 08-OCT-2004
LOCUS Callithrix jacchus clone CH259-42F14, WORKING DRAFT SEQUENCE, 2
DEFINITION ordered pieces.
AC151849
AC151849.1 GI:53983836
VERSION HTG; HTGS PHASE2; HTGS DRAFT.
KEYWORDS Callithrix jacchus (white-tufted-ear marmoset)
SOURCE Callithrix jacchus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
Callithrix.
REFERENCE 1 (bases 1 to 186028)
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186028)
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2004) Genome Sciences, Lawrence Berkeley National
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT
Sequence Produced by Berkeley PGA
Web site: <http://pga.lbl.gov>
Center Code: PGABERK
Center Project Name: J108-42F14
Bac Clone Name: CH259-42F14

This sequence has been compared to sequences of other species
using VISTA (<http://www-gsd.lbl.gov/VISTA>). The results can be
viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=navalue=PPARG

The order-orientation of the draft sequence was accomplished by using:
Avid (<http://baboon.math.berkeley.edu/mavid>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

* NOTE: This is a "working draft" sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and * the accession number will be preserved.

* 1 26679: contig of 26679 bp in length

* 26680 26779: gap of unknown length

* 26780 186028: contig of 159249 bp in length.

Location/Qualifiers

1. 186028

/organism="Callithrix jacchus"

/mol_type="genomic DNA"

/db_xref="taxon:9483"

/clone="CH259-42F14"

FEATURES

source

ORIGIN

Query Match 72.1%; Score 362.8; DB 2; Length 186028;

Best Local Similarity 91.1%; Pred. No. 2.5e-75;

Matches 468; Conservative 0; Mismatches 27; Indels 19; Gaps 7;

```

Qy 1 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAAGTTTAAACT 60
Db 84782 TAGAATAATACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAAGTTTAAACT 84723

Qy 61 ATGATACATATTTGAATTCGAATTTTCTTCAGATAATGTGATAGATTAGAGATT 120
Db 84722 ATGGATACATATTTGAATTCGAATTTTCTTCAGATAATGTGATAGATTAGAGATT 84670

Qy 121 CAACGAGGTAGACACCGCAAGAAACCTTTGCCCAATAGCTTTCTGGTATTTTCATAA 180
Db 84669 CAACGAGGTAGACACCGCAAGAAACCTTTGCCCAATAGCTTTCTGGTATTTTCATAA 84610

Qy 181 GCAAGAGATTTT-AAAGTTTTCATTTAAGAGCCCAT---TGTGAATTATACACCAATAAA 236
Db 84609 GCGAGAGATTTTAAAGTTTTCATTTAAGCAATTTGATGTAATTTTATGACAAAT-AAA 84551

Qy 237 AATGCAAGTGAATTTGAACAGTCTTCTCTGATTAATTTCTAAATACAGTACGTTCAAG 296
Db 84550 AATGCAAGGCGATATTGAACAGTCTCCGCTCTTATAATTTCTAAATACGTTACGTTCAAG 84491

Qy 297 CCCCTCAGACACTGAACATGT-GGTCCCGCGAGACAGTGTGGCAATATTA---TC 352
Db 84490 CCCCTCAGACACTCAGCATGTGGGTCTCCTGGGAGACAGTGTGGCAATATTTATTTTC 84431

Qy 353 CCTGTAATGTACCAAGTCTTCCAGAGCAGTGTG---AATCATATGACACCACTTTTGTCA 409
Db 84430 CCTGTAATGTACCAAGTCTTCCAAAGCAGTGTGAACCAATTTATGACACCAATTTGTCA 84371

Qy 410 CAGCTGGCTCTTATAGACAGTGGCAGCAATTTCAAGCCAGTCTTCTGTGTTTATT 469
Db 84370 GAGCTGGCTCTTATAGACAGTGGCAGCAATTTCAAGCCAGTCTTCTGTGTTTATT 84311

Qy 470 CCCATCTCTCCCAATATTTTGGAACTGATGTCT 503
Db 84310 CCCATCTCTCCCAATATTTTGGAACTGATGTCT 84277

```

RESULT 10

AC136055/c

LOCUS

DEFINITION

***, 62 unordered pieces.

AC136055

AC136055.1 GI:24417926

HTG; HTGS PHASE1.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 191540)

Muzny, D. Marie, Metzker, M. Lee, Abramzon, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Ayalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flegg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgiev, G. E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenshewa, L., Loulsegue, H., Lozano, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B.,

Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,

Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,

Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,

Newton, N., Nguyen, N., Norris, S., Nwakoeleneh, O., Okwuonu, G.,

Olarpunagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,

Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,

Popovic, D., Primus, E., Pu, L. L., Puazo, M., Quiroz, J., Rachlin, E.,

Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,

Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,

Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,

Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,

Sisson, I., Sitter, C. D., Smaje, D., Sneed, A., Sodergren, E.,

Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,

Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,

Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D.,

Warren, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,

Waldron, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,

Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,

Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,

Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,

Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 191540)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (29-OCT-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM


```
Qy 234 AAAAATGCAAGTGGATATGAACAGTCTCTCTCTCTGATAATCTTAAATACAGTACAGTTC 293
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69190 TTTAAAGCAATCAATATTAAACAGTCTCTGCTCTGGTAAATCCAAATACGTACAGTTC 69131
Qy 294 AGGCCCTCAGACAGACATGAACATGTGGTCAACCGGCGAGACAGTGTGGCAATATTATCC 353
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69130 ACACCCCTCACAACACACTGAATGTGTGGGTCACTGGCGGAGACAGTGTAGCAACGTTTTCC 69071
Qy 354 CTGTAATGTACCAAGTCTTTGCCA---GAGCAGTGAACATTATGACACAACTTTTGTGCAC 410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69070 TTGTAATGTACCAAGTCTTTGCCAAGAAGACAGACAGCAATTAATGACACACCAATTTGTGCAC 69011
Qy 411 AGCTGGCTCTTAAT-AGGACAGTGCACGCCCAATTAAGCCAGCCAGTCTTTCTGTGTTATT 469
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69010 AGCTGGCTCTCAATCAGGACAGTGCACGCCAATTCAGGCGCTGATCCCTTCTGTGTTATT 68951
Qy 470 CCCATCTCTCCAAATATTGGAAACTGATGCT 503
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68950 CCCACCTCTCCAAATATTGGAAACTGCTCT 68917
```

RESULT 11

```
AC091418/c
LOCUS AC091418 245724 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-2G6, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC091418
VERSION AC091418.5 GI:24941354
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 245724)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Dunthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,B., Liu,W., Loulsged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newtonson,J., Newtonson,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Petters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanji,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlarczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
```

```

TITLE Direct Submission
REFERENCE 2 (bases 1 to 245724)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 245724)
Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:22855703.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUQY
Center clone name: CH230-2G6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 220652 bases at least Q40
Consensus quality: 223895 bases at least Q30
Consensus quality: 225929 bases at least Q20
Estimated insert size: 229375; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
the accession number will be preserved.
* 1 245724: contig of 245724 bp in length.
FEATURES
source
1..245724
/organism="Rattus norvegicus"
/db_type="genomic DNA"
/mol_xref="taxon:10116"
/clone="CH230-2G6"
1..1001
/notes="wgs contig"
misc_feature
243338..245724
/notes="wgs contig"
ORIGIN
Query Match 43.7%; Score 219.6; DB 2; Length 245724;
Best Local Similarity 75.3%; Pred. No. 1.3e-41;
Matches 342; Conservative 0; Mismatches 99; Indels 13; Gaps 5;
Qy 57 AACTATCGATACATATTGTAATTCCTTCTCAGATTAATGTGATTAGATTAGA 116
```

Db 15463 AAGATATGTATATATTTGAATACAGAGTATTTCTTCAGATAATGTGATTA -----AGA 15410
 QY 117 GATTCAACAGGAGTAGACACCGAAGAAACTTTTGCCCAATAAGCTT-TCTGTATTT 175
 Db 15409 ATTTCACACAGGATGGTACTTAAAGAAATATGTTGGCAATAGTTATCTAGTGTTT 15350
 QY 176 CATAGCAGAGATTTAAAGTTTTCATTTAAGAA--GCCATTGTGGAATTATACACAAATA 233
 Db 15349 CATAACTTTAAAGATTAAAGTTTCTATTTAAAAAGCCACTGGTGTATTTACTGCAA 15290
 QY 234 AAAATGCAAGTGGATTTAGACAGTCTTCTCTGTAATTTCTAAATACAGTACAGTTTC 293
 Db 15289 TTAATAAGCAATCAATATTTAAACAGTCTCTGCTCTGTTGGTAATTTCCAAATACGTACAGTTTC 15230
 QY 294 ACGCCCTTCCAGACACTGTAACATGTGTGTACCGCGAGACAGTGTGGCAATATATCC 353
 Db 15229 ACACCCCTCACACACACTGAATGTGTGGTCTACTGGCGAGACAGTGTAGCAACGTTTTCC 15170
 QY 354 CTGTAATGTACCAAGTCTTTGCCA---GAGCAGTGAACATTTATGACACAACTTTTGTGCAC 410
 Db 15169 TTGTAATGTACCAAGTCTTTGCCAAGACGACAGACATTTATGACACACCACTTTTGTGCAC 15110
 QY 411 AGCTGGCTCTTAAT-AGACAGTGTCCAGCCCAATTCACGCCAGTCTTCTGTGTATTT 469
 Db 15109 AGCTGGCTCTCAATCAGGACAGTGTCCAGCCCAATTCAGGCCCTGTATCTTCTGTGTATTT 15050
 QY 470 CCCATCTCTCCCAATATTTTGAAGACTGATGTCT 503
 Db 15049 CCCACTCTCCCAATATTTGAAGACTGTGTCT 15016

RESULT 12

AC120668/c
 LOCUS AC120668 250169 bp DNA linear HTG 23-NOV-2002
 DEFINITION Rattus norvegicus clone CH230-24K10, *** SEQUENCING IN PROGRESS
 *** 3 unordered pieces.
 AC120668
 VERSION AC120668.4 GI:25188255
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 250169)
 Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,B.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,K., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorensuhowa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
 Nwakoelamh,O., Okwundu,G., Olarunpunaagoo,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
 Plopper,F., Poindexter,A., Popovic,D., Primus,B., Pu,L.-L.,
 Puazo,M., Quiroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Sivartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villanasa,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,X., Zhou,J., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished
 2 (bases 1 to 250169)

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

Rat Genome Sequencing Consortium

Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Nov 23, 2002 this sequence version replaced gi:23322270.

The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

COMMENT

Center: Baylor College of Medicine
 Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GXLO

Center clone name: CH230-24K10

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 225414 bases at least Q40

Consensus quality: 228258 bases at least Q30

Consensus quality: 229994 bases at least Q20

Estimated insert size: 226717; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

* 1 9799: contig of 9799 bp in length
* 9800 9899: gap of unknown length
* 9900 248957: contig of 239058 bp in length
* 248958 249057: gap of unknown length
* 249058 250169: contig of 1112 bp in length.

FEATURES

source
1..250169
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-24K10"
4754..5526
note="clone boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH280541"
9906..10968
/note="wgs contig"
11406..14106
/note="wgs contig"
complement(234439..235077)
/note="clone boundary
clone_end:T7
site:EcoRI
end_sequence:BH280540"
245037..246234
/note="wgs_end_extension
clone_end:T7"
247156..248957
/note="wgs_end_extension
clone_end:T7"

ORIGIN

Query Match 43.7%; Score 219.6; DB 2; Length 250169;
Best Local Similarity 75.3%; Pred. No. 1.3e-41;
Matches 342; Conservative 0; Mismatches 99; Indels 13; Gaps 5;
Qy 57 AACTATGATACATATTGTAATCCAAATTTCTTCAGATAATGTGATAGAGATTAGA 116
Db 154629 AAGAATATGTATATTGTAATACAAGAGTATCTTCAGATAAATGTGATTA-----AGA 154576
Qy 117 GATTCAACCGGGATAGACACCGAAGAACTTTGCCAAATAAGCTT-TCTGGTATTT 175
Db 154575 ATTTCAACCAAGGATGATACCTTAAGAAAATGTGGCAATGAGTTTATCTAGTGTTT 154516
Qy 176 CATAAGCAAGAGATTAAAGTTTCCATTTAAGAA--GCCATTGTGAATTATACACAATA 233
Db 154515 CATAACTTAAAGATTAAAGTTTCTATTAAAAAGCCACTGGTGTGTTTACTGCAA 154456
Qy 234 AAAAATGCAAGTGATATTGAACAGTCTCTCTCTGATATTCTTAAATACAGTACAGTTC 293
Db 154455 TTAATAAGCAATCAATATAAACAGTCTCTCTCTGGTAATTCANAATCTGTACAGTTC 154396
Qy 294 ACGCCCTCAGGACACTGAACATGTGGTCACCGGAGACAGTGTGGCAATATTATCC 353
Db 154395 ACACCCCTCAACAACACTGAATGTGGTGCACCTGGCGAGACAGTGTAGCAACGTTTCC 154336
Qy 354 CTGTAATGTACCAAGTCTTTGCCA--GAGCAGTGAACATTATATGACAACTTTTGTGAC 410
Db 154335 TTGTAATGTACCAAGTCTTTGCCAAGAAGACAGACAGCATTTATGACACACCATTTTGTGAC 154276
Qy 411 AGCTGGCTCTTAAT-AGGACAGTGCACCAATTAAGCCAGTCCCTTCTGTGTTTATT 469
Db 154275 AGCTGGCTCTCAATCAGACAGTGCACCAATTTAGCCCTGATCCCTTCTGTGTTTATT 154216
Qy 470 CCCATCTCTCCCAATATTTCGAAACTGATGTCT 503
Db 154215 CCCACCTCTCCCAATATTTCGAAACTGGTGTCT 154182

RESULT 13

RESULT 14
AY243584

S79407 682 bp DNA linear ROD 09-DEC-1995
LOCUS mPPAR gamma 2=peroxisome proliferator-activated receptor gamma
DEFINITION isoform 2 {promoter} [mice, Genomic, 682 nt].
S79407
ACCESSION S79407.1 GI:1110564
VERSION S79407.1
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 682)
REFERENCE Zhu,Y., Qi,C., Korenberg,J.R., Chen,X.N., Noya,D., Rao,M.S. and Reddy,J.K.
STRUCTURAL organization of mouse peroxisome proliferator-activated receptor gamma (mPPAR gamma) gene: alternative promoter use and different splicing yield two mPPAR gamma isoforms
Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7921-7925 (1995)
JOURNAL 95372391
MEDLINE 7644514
PUBMED
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gbbseq 170483] from the original journal article.
FEATURES
source
1..682
/organism="Mus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10095"
616..>682
880..682
/genes="mPPAR<gamma>2"
/notes="peroxisome proliferator-activated receptor gamma isoform 2"

Query Match 43.0%; Score 216.4; DB 10; Length 682;
Best Local Similarity 77.6%; Pred. No. 1.3e-40;
Matches 354; Conservative 0; Mismatches 86; Indels 16; Gaps 7;
Qy 57 AACTATGATACATATTGTAATCCAAATTTCTTCAGATAATGTGATAGAGATTAGA 116
Db 108 AAGAATGTGTATATTGTTAGTACAAGAAATATCTTCAGAT-GTGTGATTAG-----GA 160
Qy 117 GATTCAACCGGGATAGACACCGAAGAACTTTGCCAAATAAGCTT-TCTGGTATTT 175
Db 161 GTTTCACCAAGATAAATACTTAAGAAAACTTTGCCAAATACGTTTATCTGTGTTT 220
Qy 176 CATAAGCAAGAGATTAAAGTTTCCATTTAAGAAAGCCA---TTGTGAATTATACACAAT 232
Db 221 CATAACTTAGAGATTAAAGTTTCTATTTTAAAAGCCACTGGTGTGTTTACTGCAAT 280
Qy 233 AAAAATGCAAGTGATATTGAACAGTCTCTCTCTGATATTCTTAAATACAGTACAGTT 292
Db 281 TTTAAAAGCAATCAATATTGAACATCTCTCTCTGGTAATTCCAACTACTGTACAGTT 340
Qy 293 CACGCCCTCAGGACACTGAACATGT-GGTCAACCGGCGAGACAGTGTGGCAATATTAT 351
Db 341 CACGCCCTCAGCAACAGTGAATGTGTGGGTCACTGGCGAGACAAATGTAGCAACGTTT 400
Qy 352 CCCTGTAATGTACCAAGTCTTTGCCA---GAGCAGTGAACATTATGACACACTTTTGTGTC 408
Db 401 CCTTGTAAATGTACCAAGTCTTTGCCAAGACAGACAGCATTTATGACACCACTTTTGTGTC 460
Qy 409 ACAGCTGGCTCTTAAT-AGGACAGTGCACCAATTCAGCCAGTCTCTTCTGTGTTTA 467
Db 461 ACAATGGCTCTCAGTCAGACAGTGCAGCAATTCAGGCCCTGATCTTCTGTGTTTA 520
Qy 468 TTCCCATCTCTCCCAATATTTCGAAACTGATGTCT 503
Db 521 TTCCCATCTCTCCCAATATTTCGAAACTGGTGTCT 556

LOCUS AY243584 1159 bp DNA linear ROD 01-JUL-2003
DEFINITION Mus musculus strain 129S1/SvImJ peroxisome proliferator-activated
receptor gamma transcript 2 (Pparg) gene, promoter region.
ACCESSION AY243584
VERSION AY243584.1 GI:32395962
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1159)
AUTHORS Lyons, M.A., Wittenburg, H., Li, R., Walsh, K.A., Leonard, M.R.,
Churchill, G.A., Carey, M.C. and Paigen, B.
TITLE Lith13 Encompasses Lrpap1, Which is a Likely Genetic Determinant of
Murine Cholesterol Gallstone Formation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1159)
AUTHORS Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2003) The Jackson Laboratory, 600 Main Street,
Bar Harbor, ME 04609, USA
FEATURES
source
1. .1159
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129S1/SvImJ"
/db_xref="taxon:10090"
/chromosome="6"
/map="52.7 CM"
1. .>1159
/gene="Pparg"
/note="peroxisome proliferator-activated receptor gamma
transcript 2"
1. .1159
/gene="Pparg"
gene
promoter
ORIGIN
Query Match 43.0%; Score 216.4; DB 10; Length 1159;
Best Local Similarity 77.6%; Pred. No. 1.2e-40;
Matches 354; Conservative 0; Mismatches 86; Indels 16; Gaps 7;
Qy 57 AACTATGATACATATTGAATTCCTTCCAAATTTCTTCAGATAATGTGATTAGAGATTAGA 116
Db 633 AAGAATGTGTATATGTTTGAGTACAGAAATATCTTCAGAT-GTGTGATTAG-----GA 685
Qy 117 GATTCAACACGGGATAGACACCGAAGAAACTTTGCCAAATAAGCTT-TCTGGTATTT 175
Db 686 GTTTCACCAAGATAATATCTTAAGAAAACTTTGGCCAAATACGTTTATCTGGTGT 745
Qy 176 CATAAGCAAGAGATTTAAGTTTTCATTTAAGAAGCCA---TTGTGAATTATACAAAT 232
Db 746 CATAACTTAGAGATTAAGTTTCTATTTTAAAGCCACTGGTGTGATTATTTACTGCAAT 805
Qy 233 AAAAAATGCAGTGGATATTAAGACAGTCTCTCTGATTAATTTCTTAATACAGTACAGTT 292
Db 806 TTTAAAAAGCAATCAATATTAAGCAATCTCTGCTCTGTTAAATCCAACTACTGTACAGTT 865
Qy 293 CACGCCCTCACGAGACACTGAACATGT-GGTCAACGGCGAGACAGTGTGGCAATATTAT 351
Db 866 CACGCCCTCACAGACAGTGAATGTGTGGGTCACTGGCGAGACATGTAGCAACGTTT 925
Qy 352 CCCTGTAATGTACCAAGCTTTGCCA---GAGCAGTGAACATTTATGACACAACTTTTGTGTC 408
Db 926 CCTTGTAAATGTACCAAGCTTTGCCAAGCAGCAGCAGCATTATGACACCAATTTTGTGTC 985
Qy 409 ACAGCTGGCTCTTAAT-AGGACAGTGCAGCAATTCAGGCCAGTCTCTTCTGTGTTTA 467
Db 986 ACAACTGGCTCTCAGTCAGGACAGTGCAGCAATTCAGGCCAGTCTCTTCTGTGTTTA 1045
Qy 468 TTCCCATCTCTCCCAAAATTTTGGAACTGATGCT 503
Db 1046 TTCCCACTCTCCCAAAATTTTGGAACTGATGCT 1081

RESULT 15
AY236531 1160 bp DNA linear ROD 01-JUL-2003
LOCUS AY236531
DEFINITION Mus musculus strain DBA/2J peroxisome proliferator-activated
receptor gamma transcript 2 (Pparg) gene, promoter region.
ACCESSION AY236531
VERSION AY236531.1 GI:32395932
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1160)
AUTHORS Lyons, M.A., Wittenburg, H., Li, R., Walsh, K.A., Leonard, M.R.,
Korstanje, R., Churchill, G.A., Carey, M.C. and Paigen, B.
TITLE Lith6 Encompasses Pparggamma and SLC21A1 Which Are Likely Genetic
Determinants of Murine Cholesterol Gallstone Formation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1160)
AUTHORS Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2003) The Jackson Laboratory, 600 Main Street,
Bar Harbor, ME 04609, USA
FEATURES
source
1. .1160
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="DBA/2J"
/db_xref="taxon:10090"
/chromosome="6"
/map="52.7 CM"
1. .>1160
/gene="Pparg"
/note="peroxisome proliferator-activated receptor gamma
transcript 2"
1. .1160
/gene="Pparg"
promoter
ORIGIN
Query Match 43.0%; Score 216.4; DB 10; Length 1160;
Best Local Similarity 77.6%; Pred. No. 1.2e-40;
Matches 354; Conservative 0; Mismatches 86; Indels 16; Gaps 7;
Qy 57 AACTATGATACATATTGAATTCCTTCCAAATTTCTTCAGATAATGTGATTAGAGATTAGA 116
Db 634 AAGAATGTGTATATGTTTGAGTACAGAAATATCTTCAGAT-GTGTGATTAG-----GA 686
Qy 117 GATTCAACACGGGATAGACACCGAAGAAACTTTGCCAAATAAGCTT-TCTGGTATTT 175
Db 687 GTTTCACCAAGATAATATCTTAAGAAAACTTTGGCCAAATACGTTTATCTGGTGT 746
Qy 176 CATAAGCAAGAGATTTAAGTTTTCATTTAAGAAGCCA---TTGTGAATTATACAAAT 232
Db 747 CATAACTTAGAGATTAAGTTTCTATTTTAAAGCCACTGGTGTGATTATTTACTGCAAT 806
Qy 233 AAAAAATGCAGTGGATATTTGAACAGTCTCTCTGATTAATTTCTTAATACAGTACAGTT 292
Db 807 TTTAAAAAGCAATCAATATTTGAACAACTCTGCTCTGTTAAATCCAACTACTGTACAGTT 866
Qy 293 CACGCCCTCACGAGACACTGAACATGT-GGTCAACGGCGAGACAGTGTGGCAATATTAT 351
Db 867 CACGCCCTCACAGACAGTGAATGTGTGGGTCACTGGCGAGACAAATGTAGCAACGTTT 926
Qy 352 CCCTGTAATGTACCAAGCTTTGCCA---GAGCAGTGAACATTTATGACACAACTTTTGTGTC 408
Db 927 CCTTGTAAATGTACCAAGCTTTGCCAAGCAGCAGCAGCATTATGACACCAATTTTGTGTC 986
Qy 409 ACAGCTGGCTCTTAAT-AGGACAGTGCAGCAATTCAGGCCAGTCTCTTCTGTGTTTA 467
Db 987 ACAACTGGCTCTCAGTCAGGACAGTGCAGCAATTCAGGCCAGTCTCTTCTGTGTTTA 1046
Qy 468 TTCCCATCTCTCCCAAAATTTTGGAACTGATGCT 503

Db
1047 TTCCCACCTCTCCCAAATATTTGAAAACTGGTGCT 1082

Search completed: November 9, 2005, 02:15:43
Job time : 2644.94 secs

THIS PAGE BLANK (USPTO)

THE POLYMER BANK (USTP)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:46:48 ; Search time 407.412 Seconds
(without alignments)
7308.644 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320

Perfect score: 503

Sequence: 1 tagagtaagtaaccttagaa.....atatattgaaactgatgtct 503

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	503	100.0	2045	2	AX19033 Human PPA
2	455.2	90.5	1100	5	Aaf23651 Human PPA
c	455.2	90.5	85873	10	Adh10008 Human chr
4	455.2	90.5	158417	13	AdS36461 Human aut
c	455.2	90.5	166043	12	AdL08127 Human gen
6	216.4	43.0	1184	10	ACC79611 Mouse Per
7	52	10.3	177	2	AX19064 Human PPA
8	45.2	9.0	200000	12	Ado47191 DNA seque
c	44.2	8.8	3162	6	AbQ72637 Human MDD
c	43.6	8.7	6215	6	AbL33191 Human imm
c	41.8	8.3	99916	6	AdI03931 Human enz
c	41.8	8.3	116297	12	AdQ97587 Human can
c	41.2	8.2	32392	6	AbL56203 AMEPV gen
c	40.8	8.1	2000	12	AdJ40655 Plant cDN
c	40.8	8.1	109661	12	AdQ97818 Human can
c	40.2	8.0	376	4	AaH93354 Human chr
c	40.2	8.0	2000	8	AdA71938 Rice gene
c	40	8.0	8056	8	AbZ10100 Haematopo
c	39.4	7.8	6191	6	AbL33216 Human imm
20	39.4	7.8	6191	6	AbK31306 Signal tr

21	39.4	7.8	6191	6	ABL70281	ABL70281 Chemical	
22	39.4	7.8	6191	6	ABN80160	ABN80160 Human che	
c	39.2	7.8	6308	6	ABL33470	ABL33470 Human imm	
c	39	7.8	6565	4	AA846466	AA846466 Tumour su	
c	39	7.8	6565	6	ABK31327	ABk31327 Signal tr	
c	39	7.8	37515	6	ABQ66998	ABq66998 Human ang	
27	38.8	7.7	1696	12	ADI43067	ADI43067 Plant tra	
28	38.8	7.7	1696	12	ADO03207	ADO03207 Corn orth	
c	39	7.8	8056	8	ABZ10246	ABz10246 Haematopo	
c	38.8	7.7	10945	6	ABK28354	ABk28354 DNA trans	
c	31	38.8	7.7	19634	8	ABZ10162	ABz10162 Haematopo
c	32	38.8	7.7	19634	8	ABZ10016	ABZ10016 Haematopo
c	33	38.8	7.7	19634	13	ADS89398	ADs89398 Oligonuc
c	34	38.8	7.7	19634	13	ADS89672	ADs89672 Oligonuc
c	35	38.4	7.6	2258	4	ABL21792	ABl21792 Drosophil
c	36	38.2	7.6	4152	6	ABL55640	ABl55640 AmEPV ABC
c	37	38.2	7.6	50000	6	ABL56201	ABl56201 AmEPV gen
c	38	38.2	7.6	96589	9	ADA02708	Ada02708 Human ZFH
c	39	38.2	7.6	96589	10	ADB72446	ADb72446 Human ZFH
c	40	38.2	7.6	96589	10	ADE95956	ADe95956 Human ZFH
c	41	38	7.6	2035	13	ADT05376	ADt05376 Haemophil
c	42	38	7.6	349980	13	ADT05649	ADt05649 Haemophil
43	37.8	7.5	2142	2	AAx84309	Aax84309 Stealth v	
44	37.8	7.5	2144	3	AAZ36928	Aaz36928 Nucleotid	
c	45	37.8	7.5	5274	11	ACN92690	ACn92690 Breast ca

ALIGNMENTS

RESULT 1

AX19033

ID AX19033 standard; DNA; 2045 BP.

XX AX19033;

XX AX19033;

DT 13-MAY-1999 (first entry)

XX Human PPAR-gamma-2 proximal promoter, exon B and intron B.

DE Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;

KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;

KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;

KW insulin resistance; non-insulin-dependent diabetes mellitus;

KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;

KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.

OS Homo sapiens.

XX WO9905161-A1.

XX 04-FEB-1999.

XX 24-JUL-1998; 98WO-US015411.

XX 25-JUL-1997; 97US-0053692P.

XX (LIGA-) LIGAND PHARM INC.

XX (INSP) INST PASTEUR.

XX Briggs MR, Saladin RS, Auwerx J, Fajas L;

XX WPI; 1999-142844/12.

XX Newly isolated nucleic acid comprising a control region of a human

XX peroxisome proliferator activated receptor (PPAR) gamma gene - useful for

XX identifying modulators that are useful in treating diseases associated

XX with abnormal levels of human PPAR-gamma gene expression.

XX Claim 10; Page 80; 102pp; English.

XX The present invention describes an isolated, purified or enriched nucleic

XX acid comprising a control region of a human peroxisome proliferator

activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful for screening for agents capable of modulating the expression of a human PPAR-gamma gene. These agents (modulators) form pharmaceutical compositions that are useful for treating diseases associated with high/low levels of human PPAR-gamma gene expression. The diseases include obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas, abnormalities associated with anti-human immunodeficiency virus (HIV) treatment, insulin resistance, non-insulin-dependent diabetes mellitus (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI) tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis and bowel cancer. The nucleic acids are useful for studying the role of the PPAR-gamma gene in various diseases and disorders. The structure of the PPAR-gamma gene enables genetic studies of PPAR-gamma mutations in humans, and evaluation of its role in disorders like insulin resistance, NIDDM, and diseases associated with altered adipose tissue function, like obesity and lipodystrophic syndromes. The nucleic acids are also useful for gene therapy and the production of transgenic animals, which are useful in screening assays. The control regions of the nucleic acids enable screening for modulators of the human PPAR-gamma gene, which are useful in designing drugs for treating disorders or diseases associated with the level of PPAR-gamma gene expression. The present sequence represents human PPAR-gamma-2 proximal promoter, exon B and intron B

SQ Sequence 2045 BP; 605 A; 422 C; 351 G; 667 T; 0 U; 0 Other;

Query Match 100.0%; Score 503; DB 2; Length 2045;
Best Local Similarity 100.0%; Pred. No. 2.5e-128;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
Db 818 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 877
Qy 61 ATGATACATATTTGAATTCCTCAATTTTCTTCAGATATGTTAGAGATTAGAGATT 120
Db 878 ATGATACATATTTGAATTCCTCAATTTTCTTCAGATATGTTAGAGATTAGAGATT 937
Qy 121 CAACAGGGATAGACACCGCAAGAAACTTTGCCCAATAAGCTTTCTGGTATTTCATAA 180
Db 938 CAACAGGGATAGACACCGCAAGAAACTTTGCCCAATAAGCTTTCTGGTATTTCATAA 997
Qy 181 GCAAGAGATTAAAGTTTCCATTTTAAAGAGCCATTGTAATATACCAATAAAAAATG 240
Db 998 GCAAGAGATTAAAGTTTCCATTTTAAAGAGCCATTGTAATATACCAATAAAAAATG 1057
Qy 241 CAAGTGGATATTGACAGTCTCTTCTGTGATAATCTTAATACAGTACAGTTTCAGGCC 300
Db 1058 CAAGTGGATATTGACAGTCTCTTCTGTGATAATCTTAATACAGTACAGTTTCAGGCC 1117
Qy 301 TCACGAGACACTGAACATGTGGTCAACCGCGAGACAGTGGCCAAATATATCCCTGTAA 360
Db 1118 TCACGAGACACTGAACATGTGGTCAACCGCGAGACAGTGGCCAAATATATCCCTGTAA 1177
Qy 361 GTACAAGTCTTGCAGAGAGTGAACATTAATGACACAACTTTTGTGACAGTGGCTCC 420
Db 1178 GTACAAGTCTTGCAGAGAGTGAACATTAATGACACAACTTTTGTGACAGTGGCTCC 1237
Qy 421 TAATAGGACAGTGGCCCAATTCAGCCAGTCTCTTCTGTGTTTATTCCTCATCTCTCC 480
Db 1238 TAATAGGACAGTGGCCCAATTCAGCCAGTCTCTTCTGTGTTTATTCCTCATCTCTCC 480
Qy 481 CAAATATTGGAAACTGTATGTCT 503
Db 1298 CAAATATTGGAAACTGTATGTCT 1320

RESULT 2

AAF23651

ID AAF23651 standard; cDNA; 1100 BP.

XX AC

XX AAF23651;

XX DT

27-MAR-2001 (first entry)

XX Human PPARgamma N-terminal sequence coding sequence.
DS Cytostatic; antiinflammatory; antisense oligonucleotide; PPARgamma;
KW peroxisome proliferator-activated receptor gamma; transcription factor;
KW nuclear hormone receptor; human; infection; inflammation; tumour; ss.
XX Homo sapiens.
XX OS
XX US6159734-A.
XX 12-DEC-2000.
XX 18-JAN-2000; 2000US-00484345.
XX 18-JAN-2000; 2000US-00484345.
XX (ISIS-) ISIS PHARM INC.
XX McKay R, Baker BF, Borchers AH;
XX WPI; 2001-070112/08.
XX P-PSDB; AAB59840.
XX Novel antisense compounds capable of modulating expression of peroxisome
XX proliferator-activated receptor gamma useful for diagnosis, prophylaxis
XX and treatment of diseases associated with expression of the receptor.
XX Example 15; Col 49-52; 40pp; English.

Peroxisome proliferator-activated receptors (PPARs) are members of the nuclear hormone receptor subfamily of transcription factors. The present invention relates to antisense oligonucleotides (see AAF23652-P23731), targeted to a nucleic acid molecule encoding human PPARgamma, which specifically hybridises with and inhibits the expression of human PPARgamma. The present sequence is the coding sequence for the N-terminal sequence of human PPARgamma. The PPARgamma antisense oligonucleotides of the present invention can be used in the diagnosis and treatment of diseases associated with the expression of PPARgamma, e.g. to prevent or delay infection, inflammation or tumour formation

SQ Sequence 1100 BP; 334 A; 215 C; 195 G; 356 T; 0 U; 0 Other;

Query Match 90.5%; Score 455.2; DB 5; Length 1100;
Best Local Similarity 96.8%; Pred. No. 3.1e-115;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
Db 239 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT 298
Qy 61 ATGGATACATATTTGAATTCCTCAATTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
Db 299 ATGGATACATATTTGAATTCCTCAATTTTCTTCAGATAATGTGATTAGAGATT 351
Qy 121 CAACAGGGATAGACACCGCAAGAAACTTTGCCCAATAAGCTTTCTGGTATTTCATAA 180
Db 352 CAACAGGGATAGACACCGCAAGAAACTTTGCCCAATAAGCTTTCTGGTATTTCATAA 411
Qy 181 GCAAGAGATTAAAGTTTTCATTTAAGAGCCATTGTGAATTATACAACTAATAAATG 240
Db 412 GCAAGAGATTAAAGTTTTCATTTAAGAGCCATTGTGAATTATACAACTAATAAATG 471
Qy 241 CAAGTGGATATTGAACAGTCTCTTCTGTGATAATCTTAATACAGTACAGTTCACGCC 300
Db 472 CAAGTGGATATTGAACAGTCTCTTCTGTGATAATCTTAATACAGTACAGTTCACGCC 531
Qy 301 TCACGAGACACTGAACATGT-GGTCAACCGCGAGACAGTGTGGCAATATTATTCCTGTAA 359
Db 532 TCACGAGACACTGAACATGTGGGTCAACCGCGAGACAGTGTGGCAATATTATTCCTGTAA 591
Qy 360 TGTAACCAAGTCTTGCCAGAGCAGTGAACATATATGACACCACTTTTCTCAGCTGGCTC 419

Db 592 TGTACCAAGCTTGTGCAAAAGCAGTGAACATTTATGACACAACTTTTGTGCAGCTGGCTC 651
Qy 420 CTAATAGGACAGTGCAGCCAAATTCAGCCAGTCTTTCTGTGTTTATCCCATCTCTC 479
Db 652 CTAATAGGACAGTGCAGCCAAATTCAGCCAGTCTTTCTGTGTTTATCCCATCTCTC 711
Qy 480 CCAATATTTGGAACAGTGTCT 503
Db 712 CCAATATTTGGAACAGTGTCT 735

RESULT 3
ADH10008/c
ID ADH10008 standard; DNA; 85873 BP.
XX AC
XX ADH10008;
XX
DT 11-MAR-2004 (first entry)
XX
XX Human chromosome 3p25 DNA fragment.
XX
XX ds; human; chromosome 2; hyperplasia; tumour; thyroid; cancer;
XX break point; chromosomal band 2p21-22; DRIP; cytostatic; thyromimetic;
XX chromosome 7p15.
XX
XX Synthetic.
OS
XX Homo sapiens.
XX
XX WO2003093310-A1.
XX
XX 13-NOV-2003.
XX
XX 02-MAY-2003; 2003WO-EP004642.
XX
XX 01-MAY-2002; 2002DE-01019413.
XX
XX 14-SEP-2002; 2002DE-01042705.
XX
XX (UYBR-) UNIV BREMEN.
XX
XX Bullerdiel J;
XX
XX WPI; 2003-854480/79.
XX
XX New nucleic acid from human chromosome 2, useful for treatment and
XX diagnosis of thyroid disease, especially cancer, also related
XX polypeptides and modulators.
XX
XX Claim 1; SEQ ID NO 3; 461bp; German.

XX This invention describes a novel nucleic acid from human chromosome 2
XX that shows altered expression in hyperplasia and/or tumours, especially
XX of the thyroid. The invention describes a method for preparing nucleic
XX acid that can be detected in thyroid cancers where these contain an
XX aberration of the break point in chromosomal band 2p21-22. The invention
XX also describes constructs comprising a fragment of the human DRIP gene
XX and its splice variants, one of which contains all of exons 1-38 while
XX the other lacks exons 27 and 28. Some tumours contain a fusion protein of
XX DRIP, including a small segment from chromosome 3. The products of the
XX invention have cytostatic and thyromimetic activity and are used for the
XX inhibition of nucleic acid expression by antisense, ribozyme or RNA
XX interference (RNAi) methods. This sequence represents a fragment of human
XX chromosome 3p15 near the region encoding the DRIP protein described in
XX the disclosure of the invention.
XX
XX Sequence 85873 BP; 27329 A; 16422 C; 15988 G; 26134 T; 0 U; 0 Other;
SQ

Query Match 90.5%; Score 455.2; DB 10; Length 85873;
Best Local Similarity 96.8%; Pred. No. 1.3e-114;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;
Qy 1 TAGAGTAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAACGTTTAACT 60
Db 54472 TAGAGTAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGATACCAACGTTTAACT 54413

Qy 61 ATGATACATATTTGAATTTCCAAATTTTCTTCAGATAATGTGATTAGAGATT 120
Db 54412 ATGATACATATTTGAATTTCCAAATTTTCTTCAAATATGT-----GATTAGAGATT 54360
Qy 121 CAACAGGAGATAGACACGAAAGAAACTTTGGCCAAATAAGCTTTCTGGTATTTCTATAA 180
Db 54359 CAACAGGAGATAGACACGAAAGAAACTTTGGCCAAATAAGCTTTCTGGTATTTCTATAA 54300
Qy 181 GCAAGAGATTTAAGTTTTCATTTTAAAGAGCCATTTGTGAATTTATACAACAATAAAAAATG 240
Db 54299 GCAAGAGATTTAAGTTTTCATTTTAAAGAGCAATTTGTGAATTTTACAACAATAAAAAATG 54240
Qy 241 CAAGTGGATATTTGAACAGTCTCTTCTCTGTGATAATTTCTAAATACAGTTCACGCCCC 300
Db 54239 CAAGTGGATATTTGAACAGTCTCTCTCTGTGATAATTTCTAAATACAGTTCACGCCCC 54180
Qy 301 TCACAGACACTGAACATGT--GGTCACCGGCGAGACAGTGTGGCAATATTTATCCCTGTAA 359
Db 54179 TCACAAGACACTGAACATGTGTGGTTCACCGGCGAGACAGTGTGGCAATATTTATCCCTGTAA 54120
Qy 360 TGTACCAAGTCTTCCAGAGCAGTGAACATTTATGACACAACTTTTGTGCAGCTGGCTC 419
Db 54119 TGTACCAAGTCTTCCAGAGCAGTGAACATTTATGACACAACTTTTGTGCAGCTGGCTC 54060
Qy 420 CTAATAGGACAGTGCAGCCAAATTCAGCCAGTCTTTCTGTGTTTATTTCCATCTCTC 479
Db 54059 CTAATAGGACAGTGCAGCCAAATTCAGCCAGTCTTTCTGTGTTTATTTCCATCTCTC 54000
Qy 480 CCAATATTTGGAACAGTGTCT 503
Db 53999 CCAATATTTGGAACAGTGTCT 53976

RESULT 4
ADS36461
ID ADS36461 standard; DNA; 158417 BP.
XX AC
XX ADS36461;
XX
DT 16-DEC-2004 (first entry)
XX
XX Human autoimmune disease-related genomic DNA sequence - SEQ ID 1675.
DE single nucleotide polymorphism detection; SNP detection;
XX rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
XX systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
XX thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
XX glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
XX primary systemic vasculitis; ds.
XX Homo sapiens.
OS
XX WO2004083403-A2.
XX
XX 30-SEP-2004.
XX
XX 18-MAR-2004; 2004WO-US008461.
XX
XX 18-MAR-2003; 2003US-0455444P.
XX 25-APR-2003; 2003US-0465241P.
XX
XX (APPL-) APPLERA CORP.
XX
XX Cargill M, Begovich AB, Alexander HC;
XX
XX WPI; 2004-728480/71.
XX
XX New isolated nucleic acid molecule comprises at least 8 contiguous
XX nucleotides where one of the nucleotides is a single nucleotide
XX polymorphism (SNP), useful for diagnosing or treating autoimmune
XX diseases, e.g. rheumatoid arthritis.
XX

PS Claim 16; SEQ ID NO 1675; 123pp; English.

XX The invention comprises amino acid and coding sequences containing

CC genetic polymorphisms associated with an altered risk of developing an

CC autoimmune disease (e.g. rheumatoid arthritis). The invention further

CC comprises a method of identifying an individual that has an altered risk

CC of developing an autoimmune disease, comprising detecting a single

CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA

CC and protein sequences of the invention are useful for diagnosing and

CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1

CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory

CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious

CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,

CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The

CC present nucleic acid represents a human autoimmune disease-related

CC genomic DNA sequence of the invention. NOTE: The present sequence is not

XX shown in the specification, but has been retrieved from the WIPO website.

SQ Sequence 158417 BP; 47887 A; 30186 C; 31475 G; 48577 T; 0 U; 292 Other;

Query Match 90.5%; Score 455.2; DB 13; Length 158417;

Best Local Similarity 96.8%; Pred. No. 1.6e-114;

Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGTATACCAACGTTTAAACT 60

Db TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGTATACCAACGTTTAAACT 69123

Qy 61 ATGATACATATTTGAATTCCTCAATTTTCAGATTAATGTAGATAGATAGAGATT 120

Db ATGATACATATTTGAATTCCTCAATTTTCAGATTAATGTAGATAGATAGAGATT 69176

Qy 121 CAACAGGATAGACACCGAAGAACTTTGCCAATAGCTTCTGCTATTTTCATTA 180

Db CAACAGGATAGACACCGAAGAACTTTGCCAATAGCTTCTGCTATTTTCATTA 69236

Qy 181 GCAAGAGATTTAAGTTTCCATTTAAGAGCCATTTGTAATTTATACCAATAAAATG 240

Db GCAAGAGATTTAAGTTTCCATTTAAGAGCCATTTGTAATTTATACCAATAAAATG 69296

Qy 241 CAAGTGGATATGAACAGTCTCTCTCTGATTAATTTCAATPACAGTTCACGCC 300

Db CAAGTGGATATGAACAGTCTCTCTCTGATTAATTTCAATPACAGTTCACGCC 69356

Qy 301 TCACGACACCTGAACATGT-GGTACCGCGGACAGTGTGGCAATATTTATCCCTGTAA 359

Db TCACGACACCTGAACATGTGGTACCGCGGACAGTGTGGCAATATTTATCCCTGTAA 69416

Qy 360 TGTACCAAGTCTTGCCAGACAGTGAACATTTATGACACAACTTTTGTTCAGCTGGCTC 419

Db TGTACCAAGTCTTGCCAGACAGTGAACATTTATGACACAACTTTTGTTCAGCTGGCTC 69476

Qy 420 CTATAGACAGTGCAGCCAAATTCAGCCCAAGTCTTTCTGTGTTATTTCCCATCTCTC 479

Db CTATAGACAGTGCAGCCAAATTCAGCCCAAGTCTTTCTGTGTTATTTCCCATCTCTC 69536

Qy 480 CCAATATTTTGGAACTGATGCT 503

Db CCAATATTTTGGAACTGATGCT 69560

RESULT 5

ADL08127/c

ID ADL08127 standard; DNA; 166043 BP.

XX AC ADL08127;

XX DT 20-MAY-2004 (first entry)

XX DE Human gene associated with low HDL-C PPARG.

XX KW Human; ds; SNP; single nucleotide polymorphism;

KW high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;

KW coronary artery disease; gene.

XX Homo sapiens.

XX Key Location/Qualifiers

FT replace(145136,C)

FT /*tag= a

FT /standard_name= "Single nucleotide polymorphism"

XX US2004043389-A1.

XX 04-MAR-2004.

XX 04-SEP-2002; 2002US-00235192.

XX 04-SEP-2002; 2002US-00235192.

XX (VITI-) VITIVITY INC.

XX McCarthy J;

XX WPI; 2004-214170/20.

XX Determining whether a subject has, or is at risk of developing, an

PT abnormally low high density lipoprotein-C (HDL-C) level comprises

PT detecting an allelic variant of a polymorphic region from any of a set of

PT 27 genes.

XX Disclosure; SEQ ID NO 46; 37pp; English.

PS The invention relates to determining whether a subject has, or is at risk

CC of developing, an abnormally low high density lipoprotein-C (HDL-C) level

CC comprises determining whether the subject has an allelic variant of a

CC polymorphic region from any of 27 genes (alleles listed in Table 5 of the

CC specification). Also included are determining whether a male subject has,

CC or is at risk of developing, an abnormally low HDL-C level, comprising

CC determining whether the male subject has an allelic variant of a

CC polymorphic region listed in Table 5 which is associated with abnormally

CC low HDL-C levels in males, and determining whether a female subject has,

CC or is at risk of developing, an abnormally low HDL-C level, comprising

CC determining whether the female subject has an allelic variant of a

CC polymorphic region listed in Table 5 which is associated with abnormally

CC low HDL-C levels in females. The allelic variant in determining whether a

CC subject has, or is at risk of developing, an abnormally low HDL-C level

CC is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1

CC 1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The

CC allelic variant in determining whether a male subject has, or is at risk

CC of developing, an abnormally low HDL-C level, LRPI 3 CC or CT, PAI2 4 GG,

CC 1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their

CC complements. The methods are useful for diagnosing (a predisposition to)

CC abnormally low levels of low high density lipoprotein-C (HDL-C) in a

CC subject. The methods are useful in diagnosing (a predisposition to) or

CC prognosticating diseases and disorders associated with abnormal lipid

CC levels such as vascular and metabolic diseases, e.g., coronary artery

CC disease. The present sequence is a human gene containing a SNP (single

CC nucleotide polymorphism associated with low high density lipoprotein-C

CC (HDL-C) levels.

XX Sequence 166043 BP; 48570 A; 35113 C; 34446 G; 47914 T; 0 U; 0 Other;

SQ

Query Match 90.5%; Score 455.2; DB 12; Length 166043;

Best Local Similarity 96.8%; Pred. No. 1.6e-114;

Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGTATACCAACGTTTAACT 60

Db TAGAGTAAGTACCTTAGGAATATACATTTTCAGTAGCATGCTGTATACCAACGTTTAACT 145789

Qy 61 ATGATACATATTTGAATTCCTCAATTTTCAGATTAATGTAGATAGATAGAGATT 120

Db ATGATACATATTTGAATTCCTCAATTTTCAGATTAATGTAGATAGATAGAGATT 145677

QY 121 CAACGAGGATAGACACCGAAGAAACTTTGGCCAAATAAGCTTCTGGTATTTCATAA 180
 Db 145676 CAACGAGGATAGACACCGAAGAAACTTTGGCCAAATAAGCTTCTGGTATTTCATAA 145617
 QY 181 GCAAGAGATTTAAGTTTCCATTTTAAAGAGCCATTGTGAATTATACAAATAAAATG 240
 Db 145616 GCAAGAGATTTAAGTTTCCATTTTAAAGAGCAATTGTGAAATTTACAAATAAAATG 145557
 QY 241 CAAGTGGATATTGACAGCTCTCTCTCTGTAATTTCTAAATACAGTACAGTTCACGCC 300
 Db 145556 CAAGTGGATATTGACAGCTCTCTCTCTGTAATTTCTAAATACAGTACAGTTCACGCC 145497
 QY 301 TCACGAGACACTGAACATGT-GGTCACCGGAGACAGTGTGGCAATATTATCCCTGTAA 359
 Db 145496 TCACGAGACACTGAACATGTGGTGTGTCACCGGAGACAGTGTGGCAATATTATCCCTGTAA 145437
 QY 360 TGTACCAAGTCTTGCACAGCAGTGAACATTTATGACACAACTTTTTCACAGCTGGCTC 419
 Db 145436 TGTACCAAGTCTTGCACAGCAGTGAACATTTATGACACAACTTTTTCACAGCTGGCTC 145377
 QY 420 CTAATAGGACAGTGCACGCCAATTCAGCCAGTCTTCTCTGTGTTTATTCCTCATCTCTC 479
 Db 145376 CTAATAGGACAGTGCACGCCAATTCAGCCAGTCTTCTCTGTGTTTATTCCTCATCTCTC 145317
 QY 480 CCAATATTTGGAACTGATGTCT 503
 Db 145316 CCAATATTTGGAACTGATGTCT 145293

RESULT 6

ACCT9611

ID AAX19064 standard; DNA; 1184 BP.

XX AC ACC79611;

XX DT 06-AUG-2003 (first entry)

XX DE Mouse peroxisome proliferator activated protein-gamma-2 promoter SEQ:7.
 XX KW Mouse; Stat5; fat cell precursor; doxycycline; anorectic; obesity;
 XX KW peroxisome proliferator activated protein-gamma-2; promoter; ds.
 XX OS Mus musculus.

XX PN WO2003018803-A1.

XX PD 06-MAR-2003.

XX PF 22-FEB-2002; 2002WO-JP001605.

XX PR 23-AUG-2001; 2001JP-00253574.

XX PA (HELI-) HELIX RES INST INC.

XX PI Sato H, Wakao R, Wakao H;

XX DR WPI; 2003-290072/28.

XX PT Preparation of established fat cell precursors for application in
 XX PT screening candidate compounds for controlling induction of
 XX PT differentiation into fat cells, useful in treating e.g. obesity.
 XX PS Example 5; Page 56; 66pp; Japanese.

XX CC The present invention describes established fat cell precursors with a
 XX CC DNA structure where the Stat5 gene is ligated in a state which allows its
 XX CC expression downstream of a promoter domain which can be induced by
 XX CC doxycycline. Also described: (1) screening compounds with an activity of
 XX CC controlling the induction of differentiation into fat cells comprising:
 XX CC (a) inducing the differentiation into fat cells in the fat cell
 XX CC precursors; (b) contacting a test compound with these cells after
 XX CC induction of their differentiation into fat cells; (c) measuring extent
 XX CC of the differentiation into fat cells of such cells; and (d) comparing

CC the extent of differentiation into fat cells with that observed in a
 CC control for evaluation of activity of the test compound, during which the
 CC order of steps (a) and (b) can be changed or carried out simultaneously
 CC before steps (c) and (d); (2) screened compounds inhibiting the
 CC differentiation of fat cell precursors into fat cells; and (3) screened
 CC compounds promoting the differentiation of fat cell precursors into fat
 CC cells. Stat5 has anorectic activity. The method is useful for the
 CC preparation of established fat cell precursors for application in
 CC screening candidate compounds for controlling induction of
 CC differentiation into fat cells, which are useful in treating e.g.
 CC obesity. The present sequence represents the mouse peroxisome
 CC proliferator activated protein-gamma-2 (PPAR-gamma) promoter, which is
 CC used in the exemplification of the present invention
 XX SQ Sequence 1184 BP; 362 A; 227 C; 203 G; 392 T; 0 U; 0 Other;

Query Match 43.0%; Score 216.4; DB 10; Length 1184;

Best Local Similarity 77.6%; Pred. No. 2.2e-49;

Matches 354; Conservative 0; Mismatches 86; Indels 16; Gaps 7;

QY 57 AACTATGGATACATATTGAAATTCCAAATTTTCTTCAGATAATGTGATTAGAGATTAGA 116
 Db 677 AAGATGTGTATATGTTTGAGTACAGATATTCTTCAGAT-GTGTGATTAG-----GA 729
 QY 117 GATTCAACCGAGGATAGACACCGAAGAAACTTTGCCCAAATAAGCTT-TCTGTATT 175
 Db 730 GTTTCACACCAAGATATAATCTTAAGAAAAAACTTTGGCCAAATACGTTTATCTGTGTTT 789
 QY 176 CATAGCAGAGATTTAAGTTTTCATTTAAGAGCCCA---TTGTGAATTATACACAAT 232
 Db 790 CATAACTTAGAGATTAGGTTTCTATTATTAAGAGCCCACTGGTGTGTTATTTTCTGCAAT 849
 QY 233 AAAAAATGCAAGTGGATATTGAACAGTCTCTCTCTGATAATTCTTAAATACAGTACAGTT 292
 Db 850 TTTAAAAAGCAATCAATATTGAACAATCTCTGCTCTGTAATTCCACTACTGTACAGTT 909
 QY 293 CACGCCCTCTACGAGACACTGAACATGT-GGTACCGCGGAGACAGTGTGGCAATATTAT 351
 Db 910 CACGCCCTCTACAGAACAGTGAATGTGTGGTCTCACTGGCGAGACAATGTAGCAAGTTT 969
 QY 352 CCCTGTAATGTACCAAGTCTTGCCA---GACGAGTGAACATTATGACACAACTTTTGTGTC 408
 Db 970 CCTTGAATGTACCAAGTCTTGCCAAGCAGCAGACAGCATTTATGACACCACTTTTGTGTC 1029
 QY 409 ACAGCTGGCTCCTTAAT-AGGACAGTGCAGCCCAATTCAGGCCAGTCTCTTCTGTGTTTA 467
 Db 1030 ACACTGGCTCTCAGTCAGGACAGTGCAGCCCAATTCAGGCCAGTCTTCTTCTGTGTTTA 1089
 QY 468 TTCCCATCTCTCCCAATATTGGAACACTGATGTCT 503
 Db 1090 TTCCCATCTCTCCCAATATTGGAACACTGATGTCT 1125

RESULT 7

AAX19064

ID AAX19064 standard; DNA; 177 BP.

XX AC AAX19064;

XX DT 13-MAY-1999 (first entry)

XX DE Human PPAR-gamma-2 proximal promoter.

XX KW Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
 XX KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
 XX KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
 XX KW insulin resistance; non-insulin-dependent diabetes mellitus;
 XX KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
 XX KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.

XX OS Homo sapiens.

XX PN WO9905161-A1.

KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
KW hepatotropic; antiinflammatory; antipsoriatic; cytoskeletal; anti-HIV;
KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antigout;
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200240715-A2.
XX
XX 23-MAY-2002.
XX
XX 06-SEP-2001; 2001WO-US027628.
XX
XX 05-SEP-2000; 2000US-0229747P.
XX
XX 05-SEP-2000; 2000US-0229748P.
XX
XX 05-SEP-2000; 2000US-0229749P.
XX
XX 05-SEP-2000; 2000US-0229750P.
XX
XX 05-SEP-2000; 2000US-0229751P.
XX
XX 06-SEP-2000; 2000US-0230583P.
XX
XX 06-SEP-2000; 2000US-0230585P.
XX
XX 06-SEP-2000; 2000US-0230514P.
XX
XX 06-SEP-2000; 2000US-0230515P.
XX
XX 06-SEP-2000; 2000US-0230517P.
XX
XX 06-SEP-2000; 2000US-0230518P.
XX
XX 06-SEP-2000; 2000US-0230519P.
XX
XX 06-SEP-2000; 2000US-0230595P.
XX
XX 06-SEP-2000; 2000US-0230597P.
XX
XX 06-SEP-2000; 2000US-0230598P.
XX
XX 06-SEP-2000; 2000US-0230599P.
XX
XX 06-SEP-2000; 2000US-0230610P.
XX
XX 06-SEP-2000; 2000US-0230865P.
XX
XX 06-SEP-2000; 2000US-0230988P.
XX
XX 06-SEP-2000; 2000US-0230989P.
XX
XX 07-SEP-2000; 2000US-0230951P.
XX
XX 07-SEP-2000; 2000US-0231163P.
XX
XX 07-SEP-2000; 2000US-0231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
PI Hillman JL, Jones AU, Yu JY, Wright RU, Gietzen D, Liu TF, Yap PE;
PI Dahl CR, Momiya MG, Bradley DL, Rohatgi SD, Harris B;
PI Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;
PI Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
XX
XX WPI; 2002-527544/56.
DR P-PSDB; ABP51420.
XX
XX Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
PT AIDS.
XX
XX Claim 1; Page 407-408; 618pp; English.
PS
XX The invention relates to an isolated human disease detection and
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDPT in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDPT in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an

CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDPT. (I)
CC or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDPT, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germline gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences
XX
XX SQ Sequence 3162 BP; 971 A; 580 C; 568 G; 943 T; 0 U; 0 Other;
Query Match 8.8%; Score 44.2; DB 6; Length 3162;
Best Local Similarity 46.3%; Pred. No. 0.092;
Matches 145; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
Qy 13 CTTAGGATATTAACATTTTCAGTCAGTCGCTGATACCAACGTTTAACTATGATACATAT 72
Db 2842 CTTGAAATATAAATATATCTATAAATTTCTATGTACATAGTTTAAAAAATTTGAAATTTT 2783
Qy 73 TTGAATTCCTCAATTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACACGAGATA 132
Db 2782 TCTAATATATAAGAAACCTTAAAAAAGACAGAGAAATGCTTCCCTTAACATGTTTTC 2723
Qy 133 GACACCGAAAGAAACCTTTGCGCCAAATTAAGCTTTCTGGTATTTCATAGCAAGAGATTTA 192
Db 2722 CTATCAGAAATGAACCTTTTTCAGCTAATTCACCTTCTAATTAAGTAAACACAGAGAGCTTGA 2663
Qy 193 AGTTTTCATTAAGAACCCATTGTGATTATACAACTAAATAAATCAAGTGTGATATT 252
Db 2662 AATCTAAGTTTAAAGAGATCAAAAATAGTATACAACTTAAATAATCATATGTTCTT 2603
Qy 253 GAACAGTCTCTCTCTGATAATTTCTAATAACAGTACAGTTCACGCCCCCTCAGAGACACT 312
Db 2602 GGCCTCAGAAATTTTATTTTAAAGGTTATACAGGATCAGATTATGAGACTGTGCAACCT 2543
Qy 313 GAACATGTGTCA 325
Db 2542 GGGTAATCTGTCA 2530
RESULT 10
ABL33191/c
ID ABL33191 standard; DNA; 6215 BP.
XX
XX AC ABL33191;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Human immune system associated gene SEQ ID NO: 1164.
XX
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytoskeletal; neurotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.
XX
XX OS Homo sapiens.
XX
XX WO200200928-A2.
XX
XX PD 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.

XX	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
XX	(EPIG-) EPIGENOMICS AG.
PA	
XX	Olek A, Piepenbrock C, Berlin K;
PI	WPI; 2002-130909/17.
XX	
DR	
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful for
PT	diagnosis and treatment of diseases associated with abnormal cytosine
PT	methylation.
XX	
PS	Claim 1; SEQ ID NO 1164; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention
XX	
SQ	Sequence 6215 BP; 1813 A; 169 C; 1450 G; 2783 T; 0 U; 0 Other;
	Query Match 8.7%; Score 43.6; DB 6; Length 6215;
	Best Local Similarity 58.5%; Pred. No. 0.17; Indels 0; Gaps 0;
	Matches 76; Conservative 0; Mismatches 54;
Qy	167 CTGTATTTTCATGAACGAGATTAAAGTTTTCATTTTCCATTTAAGAGCCATTGTGAATTATAC 226
Db	5118 CTTTCATTTCTTAACCAATAACCAACATTACCATTTAAAAATTTTATATTAATAACT 5059
Qy	227 AACATAAAAATGCAAGTGATATTGAACAGTCTCTTCTCTGATAATTTCTAAATACAGT 286
Db	5058 AAAAATCTCAAAAAAAAATAAACAACATTCCTCCCTTCTCTTATAACACATCTACAAA 4999
Qy	287 ACAGTTTCACG 296
Db	4998 ATTATTCACG 4989
RESULT 11	
AD103931/C	
ID	AD103931 standard; DNA; 99916 BP.
XX	AD103931;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	Human enzyme protein encoding genomic DNA.
XX	
KW	Enzyme protein; drug screening; disease diagnosis; human; gene therapy;
KW	Chromosome 6; gene; ds; glucuronyltransferase.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	CDS 2614..96915
FT	/*tag= b
FT	/product= "enzyme protein"
FT	/note= "contains introns"
FT	2614..3204
FT	/*tag= a
FT	/number= 1
FT	3205..64312
FT	/*tag= C
FT	/number= 1
FT	64313..64457
FT	/*tag= d
FT	/number= 2

Qy 222 TATACAAACATAAAAAA 238
Db 8710 AAGGAAAAAGTAACAA 8694

RESULT 12

ADQ97587/c
ID ADQ97587 standard; DNA; 116297 BP.

AC AC
XX ADQ97587;
XX 07-OCT-2004 (first entry)

DT 07-OCT-2004 (first entry)

DE Human cancer associated sequence HD10-009, SEQ ID 564.

XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.

XX Homo sapiens.

OS WO2004060304-A2.

XX 22-JUL-2004.

XX 22-DEC-2003; 2003WO-US041389.

XX 27-DEC-2002; 2002US-00330773.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-543781/52.

XX New isolated cancer associated nucleic acids comprising at least 10
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX cancers such as leukemia and lymphoma.

XX Claim 1; SEQ ID NO 564; 199pp; English.

XX The present invention relates to cancer associated sequences (ADQ97025-
XX ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 116297 BP; 33950 A; 22909 C; 23278 G; 36160 T; 0 U; 0 Other;

Query Match 8.3%; Score 41.8; DB 12; Length 116297;

Best Local Similarity 50.8%; Pred. No. 1.4;

Matches 100; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 42 TGATACCAACGTTTAAACTATGGATACATATTTGCAATTTCCAAATTTTCTTCAGATATG 101

Db 16889 TGTAACCTGACATTTGTCATGTCATCCCTTAAACTTAAGTATTAATAATAA 16830

Qy 102 TGATTAGAGATTAGAGATTCAACCCAGGATAGACACCGAAGAAACCTTTGCCCAATPAA 161

Db 16829 ATAAATAAATAAAGAAATATCCAGGAAGATCAACCCACAGACAAAGAGCTAATAA 16770

Qy 162 GCTTCTCGTATTTCAAGCAAGAGATTAAAGTTTCCATTTAAGAGCCATTTGTGAAT 221

Db 16769 AGTACATTTTCTTGCAAAAAAATAAATTTTATTTTCAATGTAATACATTAACAACAA 16710

Qy 222 TATACAAACATAAAAAA 238

Db 16709 AAGGAAAAAGTAACAA 16693

RESULT 13

ABL56203

ID ABL56203 standard; DNA; 32392 BP.

XX

AC

XX ABL56203;

DT 01-JUL-2002 (first entry)

DE AmEPV genome fragment#5.

XX AmBPV; gene therapy; viral vector; chromosome mapping; gene mapping;
XX genetic deficiency disorder; ds.

OS Amsacta moorei entomopoxvirus.

XX WO200212526-A2.

XX 14-FEB-2002.

XX 10-AUG-2001; 2001WO-US025287.

XX 10-AUG-2000; 2000US-0224479P.

XX 14-SEP-2000; 2000US-00662254.

XX (UYFL) UNIV FLORIDA.

XX Moyer RW, Li Y, Bawden AL;

XX WPI; 2002-227161/28.

XX Novel recombinant entomopox virus vector useful for delivering
XX polynucleotide encoding protein to vertebrate cell, comprises
XX polynucleotide encoding protein operably linked with heterologous
XX promoter sequence.

XX Disclosure; Page 226-242; 326pp; English.

XX The invention relates to a recombinant entomopox virus (EPV) vector,
XX comprising a polynucleotide encoding a protein operably linked with a
XX heterologous promoter sequence. The invention also concerns methods for
XX providing gene therapy for genetic deficiency disorders. Vectors of the
XX invention are useful for delivering a polynucleotide encoding a protein
XX to a vertebrate cell preferably a mammalian cell, such as a human cell.
XX The vector is introduced into the vertebrate cell by infection in a viral
XX particle, or by transfection, transduction, or injection either in vitro
XX or in vivo. The vector is useful for the delivery and expression of
XX biologically useful proteins in gene therapy protocols, and for
XX delivering large DNA segments for engineering of vertebrate cells.

XX Polynucleotides of the invention have applications in techniques such as
XX their use as insertion sites for foreign genes of interest, hybridisation
XX probes for chromosome and gene mapping, in PCR technologies, and in the
XX production of sense or antisense nucleic acids. Vectors of the invention
XX provide for stable integration and expression of heterologous DNA in host
XX cells, and are adapted for accepting large heterologous polynucleotide
XX inserts which can be delivered in an infected or transformed cell and
XX expressed in a stable fraction. The current sequence represents a
XX fragment of the genome of the genus B entomopoxvirus from amsacta moorei
XX (AmEPV)

XX Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;

Query Match 8.2%; Score 41.2; DB 6; Length 32392;

Best Local Similarity 48.7%; Pred. No. 1.3;

Matches 112; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 14 TTAGGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACTATGATACATATT 73

Db 4783 TTATACATATAACTTAATAATAATTATTAATAATAATAATAATAATAATAATAATA 4842

Qy 74 TGAATTCGAATTTTCTTCAGATAATGTTAGATAGATTAGAGATTCAACCGGATAG 133

Db 4843 ATATTTATTAATAACAAATTAAGAAATCTGCAGATTAACAAACATAGAAAATATAT 4902

Qy 134 ACACGGAAGAAACCTTTGCCCAATAGCTTTCTGGTATTTTCATAACGAGAGATTTAA 193

Db 4903 TATCTTGCAAAATTTATTAATAATAAGATTATAGTGTATTTATGATAGATTATA 4962

Search completed: November 9, 2005, 00:12:39
Job time : 411.412 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 15:16:42 ; Search time 128.167 Seconds
(without alignments)
6421.701 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320

Perfect score: 503

Sequence: 1 tagagtaagtaccttagaa.....atattggaactgatgtct 503

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	455.2	90.5	1100	3	US-09-484-345-10
2	455.2	90.5	86877	4	US-09-949-016-15491
3	455.2	90.5	86877	4	US-09-949-016-15492
4	142.4	28.3	601	4	US-09-949-016-133111
5	142.4	28.3	601	4	US-09-949-016-133114
6	41.8	8.3	99916	4	US-09-816-095-3
7	40.4	8.0	1141	4	US-09-806-708B-22
8	39.4	7.8	1055	4	US-09-806-708B-23
9	38.8	7.7	53806	4	US-09-949-016-12572
10	38.4	7.6	108310	4	US-09-949-016-16366
11	38.2	7.6	136264	4	US-09-949-016-12756
12	38.2	7.6	136265	4	US-09-949-016-13001
13	38	7.6	1317	4	US-09-248-796A-1010
14	37.6	7.5	640681	4	US-09-790-988-1
15	37	7.4	192506	4	US-09-949-016-15830
16	36.8	7.3	117838	4	US-09-949-016-17595
17	36.6	7.3	1806	4	US-09-746-359A-36
18	36.2	7.2	120609	4	US-09-949-016-13915
19	35.6	7.1	601	4	US-09-949-016-155222
20	35.4	7.0	147321	4	US-09-949-016-15450
21	35.2	7.0	2193	4	US-09-011-143-5
22	35.2	7.0	2193	4	US-09-302-495-5
23	35.2	7.0	2193	4	US-10-079-616-5
24	35.2	7.0	19124	2	US-08-487-826B-13
25	35.2	7.0	64638	4	US-09-949-016-11767
26	35.2	7.0	64639	4	US-09-949-016-13520
27	35.2	7.0	251769	4	US-09-949-016-13185

c 28	35.2	7.0	251769	4	US-09-949-016-13186	Sequence 13186, A
c 29	35.2	7.0	266748	4	US-09-949-016-13187	Sequence 13187, A
c 30	35.2	7.0	266748	4	US-09-949-016-13188	Sequence 13188, A
c 31	35	7.0	798	4	US-09-270-767-5570	Sequence 5570, Ap
c 32	35	7.0	798	4	US-09-270-767-20852	Sequence 20852, A
c 33	35	7.0	966	4	US-09-270-767-21904	Sequence 21904, A
c 34	35	7.0	966	4	US-09-270-767-436	Sequence 436, App
c 35	35	7.0	1457	4	US-09-270-767-15718	Sequence 15718, A
c 36	35	7.0	1457	4	US-09-270-767-17604	Sequence 17604, A
c 37	34.6	6.9	15861	4	US-09-949-016-14543	Sequence 14543, A
c 38	34.6	6.9	387902	4	US-09-949-016-12557	Sequence 12557, A
c 39	34.6	6.9	421883	4	US-09-949-016-155221	Sequence 155221, A
c 40	34.4	6.8	601	4	US-09-806-708B-22	Sequence 22, Appl
c 41	34.4	6.8	1141	4	US-09-248-796A-1619	Sequence 1619, Ap
c 42	34.4	6.8	1458	4	US-09-949-016-16294	Sequence 16294, A
c 43	34.4	6.8	43981	4	US-09-949-016-14429	Sequence 14429, A
c 44	34.4	6.8	96987	4	US-09-949-016-16089	Sequence 16089, A
c 45	34.4	6.8	175265	4	US-09-949-016-16089	Sequence 16089, A

ALIGNMENTS

RESULT 1
US-09-484-345-10
; Sequence 10, Application US/09484345
; Patent No. 6159734
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Alexander H. Borchers
; APPLICANT: Brenda F. Baker
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEROXISOME PROLIFERATOR-ACTIVATED RECEPTO
; FILE REFERENCE: RTS-0104
; CURRENT APPLICATION NUMBER: US/09/484,345
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (859)....(940)
US-09-484-345-10

Query Match	90.5%	Score 455.2;	DB 3;	Length 1100;
Best Local Similarity	96.8%	Pred. No. 1.8e-124;	Indels 8;	Gaps 2;
Matches 486;	Conservative 0;	Mismatches 8;		
Qy	1	TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT	60	
Db	239	TAGAGTAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACT	298	
Qy	61	ATGATACATATTTTGAATTTTCCAAATTTTCTTCAGATAATGATAGATTAGAGATT	120	
Db	299	ATGATACATATTTTGAATTTTCCAAATTTTCTTCAGATAATGATAGATTAGAGATT	351	
Qy	121	CAACCGGATAGACACCGGAAGAAACTTTGCCAAATAGCTTTCTGTATTTCATAA	180	
Db	352	CAACCGGATAGACACCGGAAGAAACTTTGCCAAATAGCTTTCTGTATTTCATAA	411	
Qy	181	GCAGAGATTAAAGTTTTCATTTTGAAGAGCATTTGAAATTTATCAACATAAAAAATG	240	
Db	412	GCAGAGATTAAAGTTTTCATTTTGAAGAGCATTTGAAATTTTATCAACATAAAAAATG	471	
Qy	241	CAAGTGATATTAAGAGTCTCTCTCTGATAATTTCTGATAATTTCTGATAATTTCTGAGTTCAGGCCCC	300	
Db	472	CAAGTGATATTAAGAGTCTCTCTCTGATAATTTCTGATAATTTCTGATAATTTCTGAGTTCAGGCCCC	531	
Qy	301	TCACGACACTGAACATGT-GGTTCACCGGAGACAGTGTGGCAATATTATTCCTCTGTA	359	
Db	532	TCACGACACTGAACATGTGGTTCACCGGAGACAGTGTGGCAATATTATTCCTCTGTA	591	


```
US-09-949-016-133111
; Sequence 133111, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133111
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-133111

Query Match      28.3%; Score 142.4; DB 4; Length 601;
Best Local Similarity 99.3%; Pred. No. 3.5e-32;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 360 TGTACCAAGTCTTCCAGAGCAGTGAACTATATGACACAACTTTTGTGCAGCTGGCTC 419
Db 1 TGTACCAAGTCTTCCAGAGCAGTGAACTATATGACACAACTTTTGTGCAGCTGGCTC 60

Qy 420 CTAATAGGACAGTGCCAGCCCAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 479
Db 61 CTAATAGGACAGTGCCAGCCCAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 120

Qy 480 CCAAATATTTGGAACTGATGCT 503
Db 121 CCAAATATTTGGAACTGATGCT 144

Query Match      28.3%; Score 142.4; DB 4; Length 601;
Best Local Similarity 99.3%; Pred. No. 3.5e-32;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 360 TGTACCAAGTCTTCCAGAGCAGTGAACTATATGACACAACTTTTGTGCAGCTGGCTC 419
Db 1 TGTACCAAGTCTTCCAGAGCAGTGAACTATATGACACAACTTTTGTGCAGCTGGCTC 60

Qy 420 CTAATAGGACAGTGCCAGCCCAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 479
Db 61 CTAATAGGACAGTGCCAGCCCAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 120

Qy 480 CCAAATATTTGGAACTGATGCT 503
Db 121 CCAAATATTTGGAACTGATGCT 144

RESULT 5
US-09-949-016-133194
; Sequence 133194, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133194
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-133194

Query Match      28.3%; Score 142.4; DB 4; Length 601;
Best Local Similarity 99.3%; Pred. No. 3.5e-32;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 360 TGTACCAAGTCTTCCAGAGCAGTGAACTATATGACACAACTTTTGTGCAGCTGGCTC 419
Db 1 TGTACCAAGTCTTCCAGAGCAGTGAACTATATGACACAACTTTTGTGCAGCTGGCTC 60
```

```
Qy 420 CTAATAGGACAGTGCCAGCCCAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 479
Db 61 CTAATAGGACAGTGCCAGCCCAATTCAGCCAGTCTTTCTGTGTTATTCCTCATCTCTC 120

Qy 480 CCAAATATTTGGAACTGATGCT 503
Db 121 CCAAATATTTGGAACTGATGCT 144

RESULT 6
US-09-816-095-3/c
; Sequence 3, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 99916
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(99916)
; OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

Query Match      8.3%; Score 41.8; DB 4; Length 99916;
Best Local Similarity 50.8%; Pred. No. 0.15;
Matches 100; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 42 TGATACCAACGTTTAAACTATGGATACATATTTGCAATTTTCCAAATTTTCTTCAGATAATG 101
Db 8890 TGTAACCTAACCTGCACATTTGTGCACATGTACCTTAAACTTAAAGTATAATAATAA 8831

Qy 102 TGATTAGAGATTAGAGATTCAACAGGATAGACACCGAAGAACTTTGCCCAATAA 161
Db 8830 ATAAATATAAATAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATA 8771

Qy 162 GCTTTCTGTGTTTTCATAAGCAAGAGATTTAAAGTTTTCATTTAAGAGCCATTGTGAAT 221
Db 8770 AGTACATTTCTTCCAAAAAATAATTTTATTTTCAATGTATATTAACAACAA 8711

Qy 222 TATACAAACAATAAAAAA 238
Db 8710 AAGGAAAAAAGTAACAA 8694

RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
```


Query Match 7.6%; Score 38.2; DB 4; Length 136264;
Best Local Similarity 49.8%; Pred. No. 2;

	Query Match	7.6%;	Score 38.2;	DB 4;	Length 136265;
	Best Local Similarity	49.8%;	Pred. No. 2;		
	Matches 126;	Conservative 0;	Mismatches 123;	Indels 4;	Gaps 1;
Qy	66	TACATATTTTGAATTCCAAATTTTCTTCAGATAAATGTGATTAGAGATTAGAGATTCAACC	125		
Db	24356	TATATAATATATATTAAATATTATATATTATATAAATATATATTAAATTATATATTATA	-- 24299		
Qy	126	AGGGATAGACACGGAAGAAAACITTTGCCRAATAAGCTTCTCGTATTTCATAAGCAAG	185		
Db	24298	--TAATATATATATTAATAATATATATATATAATATATATATATATATATATATATAT	24241		
Qy	186	AGATTTTAAGTTTTCCATTTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGT	245		
Db	24240	ATATATTAAT	24181		
Qy	246	GGATATTGACAGTCTCTCTCTGTAATTTCTAAATACAGTACAGTTTCACGCCCTCAGG	305		
Db	24180	TAATATATATATATATATAATATACTATATATATATATATATATATATATATATATATAT	24121		
Qy	306	AGACACTGGAACAT	318		

```
Db 24120 AGTAACTCCACAT 24108
|||||
RESULT 13
US-09-248-796A-1010
; Sequence 1010, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 1010
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-1010
Query Match 7.6%; Score 38; DB 4; Length 1317;
Best Local Similarity 47.8%; Pred. No. 0.34; Mismatches 120; Indels 0; Gaps 0;
Matches 110; Conservative 0;
53 TTTAACTATGATGATACATATTTGAATTCCTCAAAATTTTCTTCAGATAATGTGATTAGAT 112
|||||
624 TTTAAATAATTTTAAACAATTTGGAATTAATTAATTAGGATCATCGAAACATTTAAAAAC 683
|||||
113 TAGAGATTCACACAGGGATAGACACCGAAAGAAACCTTTGCCAAATAAGCTTTCTGGTA 172
|||||
684 TTTAGATATAAATTTAAATTTAGAACAAAAAATCCACTACCAAGAGAAATTTTAATGAA 743
|||||
173 TTTCCATAGCAGAGATTTAGATTTTCCATTTTAAAGAGCCATGCGAATTATACAACAAT 232
|||||
744 ATGGAGTATTAATTAATGATTTTAAATTTCAAAAACATTAAGAAATTTAAAAATTAAT 803
|||||
233 AAAAAATGCAAGTGGATATTTGAACAGATCTCTTCTCGATAATTTCTAAATA 282
|||||
804 TAATAACAATCGTGGTAATGGAATCATTTTTTAATTAATATGTTTATA 853
|||||
RESULT 14
US-09-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1
Query Match 7.5%; Score 37.6; DB 4; Length 640681;
Best Local Similarity 49.5%; Pred. No. 5.6;
Matches 97; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
```

```
Qy 44 ATACCAAGTTTAAACTATGATACATATTTGAATTCCTCAAAATTTTCTTCAGATAATGTG 103
|||||
Db 212263 ATAAATAAATTTCTATATATACATAAATAATAAATAATATATATTTTTTAACTATGCACT 212204
|||||
Qy 104 ATTAGAGATTAGAGATTCAACACGAGGATAGACACCGAAAGAAACCTTTGCCAAATAAGC 163
|||||
Db 212203 AATAATGATCATTTATGTTATATAAAATTAATGCAATCTAAAAAGTTTTTAAAAACAGTTC 212144
|||||
Qy 164 TTTCTGTTATTTATATAGCAGAGATTTAAGTTTCCATTTTCCATTTTACAGCCATTTGGAATTA 223
|||||
Db 212143 TTTTATATATTTTAAAAAATGAAAAATATATATATTTTTTATATAAAAAATATATATTTCTTTTA 212084
|||||
Qy 224 TACAACAATAAAAAAT 239
|||||
Db 212083 TACCTAATTATAAAT 212068
|||||
RESULT 15
US-09-949-016-15830/c
; Sequence 15830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15830
; LENGTH: 192506
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15830
Query Match 7.4%; Score 37; DB 4; Length 192506;
Best Local Similarity 52.2%; Pred. No. 5.1;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
17 GGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACTATATGATACATATTGA 76
|||||
191028 GGAAGAGACCATTTAACTCTCCAACAATAAGTGACATCAAGAAAAATGACACCAATTTGT 190969
|||||
Qy 77 ATTCCAAATTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACACGGGATAGACA 136
|||||
190968 AAAACAAATTTTTTTTCCAGTAAATTCGAGAGATACAAATACAAACACTGACAGATT 190909
|||||
Qy 137 CCAGAAAGAAACCTTTGCCAAATAAGCTTTCTCGTAT 173
|||||
190908 ACTGCTTTAAGATCTATCTCTAACAAATTTCTTTTCAAAT 190872
|||||
Search completed: November 8, 2005, 19:52:47
Job time : 131.167 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 18:06:53 ; Search time 650.857 Seconds
(without alignments)
6391.142 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320

Perfect score: 503

Sequence: 1 tagagtaagtaaccttagaa.....atatattggaactgatgtct 503

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*
- 26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	455.2	90.5	1100	24	US-10-181-176-10
c 2	455.2	90.5	165043	19	US-10-235-192A-46
3	45.2	9.0	200000	20	US-10-672-764A-31
c 4	44.2	8.8	3162	20	US-10-363-829-189
c 5	43.6	8.7	6215	16	US-10-311-455-1164

Query Match 90.5%; Score 455.2; DB 24; Length 1100;

c 6	41.8	8.3	99916	9	US-09-816-095-3	Sequence 3, Appli
c 7	41.8	8.3	99916	19	US-10-634-905-3	Sequence 3, Appli
c 8	41.2	8.2	32392	22	US-10-706-635-27	Sequence 27, Appli
c 9	40.8	8.1	2000	18	US-10-260-238-1655	Sequence 1655, Ap
c 10	40.8	8.1	59914	20	US-10-741-601-5619	Sequence 5619, Ap
c 11	40.8	8.1	101782	20	US-10-741-601-5661	Sequence 5661, Ap
c 12	40.4	8.0	1886	20	US-10-437-963-36853	Sequence 36853, A
c 13	40.2	8.0	797	19	US-10-424-599-18984	Sequence 18984, A
c 14	40	8.0	8056	21	US-10-473-126-240	Sequence 240, App
c 15	39.8	7.9	350	20	US-10-674-124A-1055	Sequence 1055, App
c 16	39.4	7.8	350	18	US-10-242-535A-53661	Sequence 53661, A
c 17	39.4	7.8	350	18	US-10-085-783B-53661	Sequence 53661, A
c 18	39.4	7.8	6191	16	US-10-311-455-1189	Sequence 1189, Ap
c 19	39.2	7.8	629	13	US-09-925-065A-861470	Sequence 861470, A
c 20	39.2	7.8	967	13	US-09-925-065A-37192	Sequence 37192, A
c 21	39.2	7.8	967	13	US-09-925-065A-37193	Sequence 37193, A
c 22	39.2	7.8	1943	13	US-09-925-065A-92097	Sequence 92097, A
c 23	39.2	7.8	6308	16	US-10-311-455-1443	Sequence 1443, Ap
c 24	39	7.8	6565	19	US-10-221-714A-188	Sequence 188, App
c 25	39	7.8	37515	20	US-10-433-793-28	Sequence 28, Appli
c 26	38.8	7.7	505	14	US-10-027-632-248390	Sequence 248390, A
c 27	38.8	7.7	505	18	US-10-027-632-248390	Sequence 248390, A
c 28	38.8	7.7	1696	18	US-10-374-780A-1530	Sequence 1530, Ap
c 29	38.8	7.7	1696	19	US-10-412-699B-1620	Sequence 1620, Ap
c 30	38.8	7.7	8056	21	US-10-473-126-386	Sequence 386, App
c 31	38.8	7.7	10945	16	US-10-240-453-228	Sequence 228, App
c 32	38.8	7.7	19634	21	US-10-473-126-156	Sequence 156, App
c 33	38.8	7.7	19634	21	US-10-473-126-302	Sequence 302, App
c 34	38.4	7.6	2258	26	US-11-097-143-29929	Sequence 29929, A
c 35	38.2	7.6	586	14	US-10-027-632-197296	Sequence 197296, A
c 36	38.2	7.6	586	18	US-10-027-632-197296	Sequence 197296, A
c 37	38.2	7.6	4152	22	US-10-706-635-9	Sequence 9, Appli
c 38	38.2	7.6	50000	19	US-10-706-635-25	Sequence 25, Appli
c 39	38.2	7.6	96589	22	US-10-052-482-214	Sequence 214, App
c 40	38	7.6	629	13	US-09-925-065A-861469	Sequence 861469, A
c 41	38	7.6	629	13	US-09-925-065A-897143	Sequence 897143, A
c 42	38	7.6	2035	24	US-10-795-159-412	Sequence 412, App
c 43	38	7.6	3426	13	US-09-925-065A-676913	Sequence 676913, A
c 44	38	7.6	3426	13	US-09-925-065A-676914	Sequence 676914, A
c 45	38	7.6	908765	24	US-10-795-159-685	Sequence 685, App

ALIGNMENTS

RESULT 1

US-10-181-176-10
; Sequence 10, Application US/10181176
; Publication No. US20050186567A1
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Alexander H. Borchers
; APPLICANT: Brenda F. Baker
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR
; TITLE OF INVENTION: GAMMA EXPRESSION
; FILE REFERENCE: RTSP-0328
; CURRENT FILING DATE: 2002-07-12
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US09/484,345
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (859)...(940)
US-10-181-176-10

```
Best Local Similarity 96.8%; Pred. No. 1.6e-108;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGCTTTAAACT 60
Db 239 TAGAGTAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGCTTTAAACT 298
Qy 61 ATGGATACATATTTGAATTTCCAAATTTTCTTCAGATAATGATGATTAGAGATTAGAGATT 120
Db 299 ATGGATACATATTTGAATTTCCAAATTTTCTTCAAATAATGT-----GATTAGAGATT 351
Qy 121 CAACAGGGATAGACACCGAAGAAAACCTTTGCCAAATAAGCTTTCTGGTATTTCATAA 180
Db 352 CAACAGGAATAGACACCGAAGAAAACCTTTGCCAAATAAGCTTTCTGGTATTTCATAA 411
Qy 181 GCAAGAGATTTAAGCTTTTCATTTTAAGNAGCCATTTGAATATATACAAATAAAAAATG 240
Db 412 GCAAGAGATTTAAGCTTTTCATTTTAAGNAGCAATTTGTGAATTTTACAAATAAAAAATG 471
Qy 241 CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTAAATACAGTACAGTTACGCCCC 300
Db 472 CAAGTGGATATTGAACAGTCTCTCTCTGATAATTTCTAAATACAGTACAGTTACGCCCC 531
Qy 301 TCACGAGACACTGAACATGT-GGTCAACGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db 532 TCACAGACACTGAACATGTGGGTCAACGGCGAGACAGTGTGGCAATATTATCCCTGTAA 591
Qy 360 TGTACCAAGCTTTCAGAGACAGTGAACATTTATGACACAACCTTTTGTCAAGTGGCTC 419
Db 592 TGTACCAAGCTTTCAGAGACAGTGAACATTTATGACACAACCTTTTGTCAAGTGGCTC 651
Qy 420 CTAATAGGACAGTGCAGCCCAATTCAGCCCAAGTCTTTCTGTGTTTATCCCATCTCTC 479
Db 652 CTAATAGGACAGTGCAGCCCAATTCAGCCCAAGTCTTTCTGTGTTTATCCCATCTCTC 711
Qy 480 CCAAATATTGGAAACTGATGTCT 503
Db 712 CCAAATATTGGAAACTGATGTCT 735

RESULT 2
US-10-235-192A-46/c
; Sequence 46, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; TITLE OF INVENTION: Methods and Compositions for Identifying
; TITLE OF INVENTION: Risk Factors for Abnormal Lipid Levels and the Diseases
; TITLE OF INVENTION: and Disorders Associated Therewith
; FILE REFERENCE: MMI-011
; CURRENT APPLICATION NUMBER: US/10/235,192A
; CURRENT FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 166043
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-192A-46

Query Match 90.5%; Score 455.2; DB 19; Length 166043;
Best Local Similarity 96.8%; Pred. No. 1.3e-107;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

Qy 1 TAGAGTAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGCTTTAAACT 60
Db 145789 TAGAGTAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAAGCTTTAAACT 145730
Qy 61 ATGGATACATATTTGAATTTCCAAATTTTCTTCAGATAATGATGATTAGAGATTAGAGATT 120
Db 145729 ATGGATACATATTTGAATTTCCAAATTTTCTTCAAATAATGT-----GATTAGAGATT 145677
Qy 121 CAACAGGGATAGACACCGAAGAAAACCTTTTGCCAAATAAGCTTTCTGGTATTTCATAA 180

Best Local Similarity 9.0%; Score 45.2; DB 20; Length 200000;
Best Local Similarity 48.1%; Pred. No. 2.6;
Matches 128; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 30 TCAGTAGCATGCTGATACCAACGTTTAAACTATGATGATACATATTTGAATTTCCAAATTTT 89
Db 139429 TGAGAACATGGAAAGCTTTCTTCTAAGATAGGATAGAAAGAAATGCCAGTCTTA 139488
Qy 90 CTTGAGATAATGTGATTAGAGATTAGAGATTCAACAGGGATAGACACCGAAGAAAACCT 149
Db 139489 ATAATCTAATCAGCATATTGATGATGATGTTTAAACCATGAAATTTGAGAAAGAAAAA 139548
Qy 150 TTGCCCAATAAGCTTTCTGTGTTTTCATAGCAAGAGATTAAAGTTTTCATTTAAGAA 209
Db 139549 TAGTAAAGGCATACAAATTTGAGGGGAATTAAGTTAAATTTGTTCTGTGTTACAGACAAT 139608
Qy 210 GCCATTGTGAATTTATACAAATAAAAAATGCAAGTGGATATTGAACAGTCTCTCTCTG 269
Db 139609 AACTTTATAAATTTCCAAAAAATAAAAAATTTGAAAAAACACGCCACTGAAACACTAACAA 139668
Qy 270 ATAATTTCTAAATACAGTACAGTTTCAAC 295
Db 139669 ACAAATTTAGTAACATTCAGTATATAC 139694

RESULT 4
US-10-363-829-189/c
```



```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(99916)
; OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

Query Match      8.3%; Score 41.8; DB 9; Length 99916;
Best Local Similarity 50.8%; Pred. No. 15;
Matches 100; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 42 TGATACCACGTTTAAACTATGGATACATATTTGCAATTTCCAAATTTTCTTCAGATAATG 101
Db 8890 TGTAACTAACCTGCACATTTGTCACATGTACCTTAAATTTTAAAGTATATAATAATAA 8831

Qy 102 TGATTAGAGATTAGAGATTCAACCAAGGATAGACACCGAAGAAACTTTGCCCAATAA 161
Db 8830 ATAAATATAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGCTAATAA 8771

Qy 162 GCTTCTGGTATTTTCAATAGCAAGAGATTTAAAGTTTTCATTTTAAAGAGCCATTGTGAAT 221
Db 8770 AGTACATTTTCTGCAAAAAAATAATTTTATTTTCAATGTAATACATTAACACAA 8711

Qy 222 TATACAAATATAAAAA 238
Db 8710 AAGGAAAAAAGTAACAA 8694

RESULT 7
US-10-634-905-3/c
; Sequence 3, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinlu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001147-DIV
; CURRENT APPLICATION NUMBER: US/10/634,905
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 99916
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(99916)
; OTHER INFORMATION: n = A,T,C or G
US-10-634-905-3

Query Match      8.3%; Score 41.8; DB 19; Length 99916;
Best Local Similarity 50.8%; Pred. No. 15;
Matches 100; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 42 TGATACCACGTTTAAACTATGGATACATATTTGCAATTTCCAAATTTTCTTCAGATAATG 101
Db 8890 TGTAACTAACCTGCACATTTGTCACATGTACCTTAAATTTTAAAGTATATAATAATAA 8831

Qy 102 TGATTAGAGATTAGAGATTCAACCAAGGATAGACACCGAAGAAACTTTGCCCAATAA 161
Db 8830 ATAAATATAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATAA 8771

Qy 162 GCTTCTGGTATTTTCAATAGCAAGAGATTTAAAGTTTTCATTTTAAAGAGCCATTGTGAAT 221
Db 8770 AGTACATTTTCTGCAAAAAAATAATTTTATTTTCAATGTAATACATTAACACAA 8711

Qy 222 TATACAAATATAAAAA 238
Db 8710 AAGGAAAAAAGTAACAA 8694
```

RESULT 8

```
US-10-706-635-27
; Sequence 27, Application US/10706635
; Publication No. US20050014263A1
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; FILE REFERENCE: US-221C1XCZ1
; CURRENT APPLICATION NUMBER: US/10/706,635
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 09/662,254
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 32392
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-27

Query Match      8.2%; Score 41.2; DB 22; Length 32392;
Best Local Similarity 48.7%; Pred. No. 13;
Matches 112; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 14 TTAGGAATATAACATTTTTCAGTAGCATGCTGATACCAACGTTTAAACTATGATACATATT 73
Db 4783 TTATACATATACTTAATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4842

Qy 74 TGAATCCAAATTTTCTTCAGATAATGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133
Db 4843 ATATTATTATTAACAAATTTAAAGAAATTTCTGCAGATTACAAAACTAGAAAAATATAT 4902

Qy 134 ACACCGAAGAAACTTTGCCAAATAAGCTTTCTGTATTTTCATAGCAAGAGATTAA 193
Db 4903 TATCTTGTCAAAATTTTATTAATAATAAGATTATAGTGTATTTTATGGAATAGATTATA 4962

Qy 194 GTTTTCATTTAAGAAGCCATTGTAATTAATAACAATAAAAAATGCAA 243
Db 4963 AAAATATAATAAAATAATTTTAAAAAATATATGTAATAAAGGATTTTAAA 5012

RESULT 9
US-10-260-238-1655/c
; Sequence 1655, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
```

; SEQ ID NO 1655
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-1655

Query Match 8.1%; Score 40.8; DB 18; Length 2000;
Best Local Similarity 48.0%; Pred. No. 5;
Matches 117; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 4 AGTAAGTACCTTAGCAATATAACATTTTCAGTAGCATCTGATACCAAGTTTAACTATG 63
Db |||||
Qy 1937 ATTTCGATGATGAACATAGAAATAATAATAGAGAGAGTGTCTTTCACTAAAT 1878
Db |||||
Qy 64 GATACATATTTGAATTCCTCAATTTTCCTCAGATAATGTAATGAGATTAGAGATCAA 123
Db |||||
Qy 1877 GACAGATTTTGTATGCTTTCATATAATTTGCATATTTATATATCATCGAGAAACTATAAG 1818
Db |||||
Qy 124 CCAGGGATAGACACGGAAGAAACTTTGCCCCAATAATAGCTTCTGGTATTTTCATAGCA 183
Db |||||
Qy 1817 CAATTGAGATGATTTGCTTTATGATATTTTAGATCAAAATATTTGCTGCTTAAAGGA 1758
Db |||||
Qy 184 AGAGATTAACTTTTCCATTTAAGAGCCATTGTGAATTATACAACTAAATAATGCAA 243
Db |||||
Qy 1757 AAAAACAATGATGATGGGACACAAAGGACAATGTGGACATAGAGCTATAAATAATTAAA 1698
Db |||||
Qy 244 GTGG 247
Db |||||
Qy 1697 ATGG 1694

RESULT 10

US-10-741-601-5619
; Sequence 5619, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5619
; LENGTH: 59914
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(59914)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5619

Query Match 8.1%; Score 40.8; DB 20; Length 59914;
Best Local Similarity 50.5%; Pred. No. 22;
Matches 99; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 76 AATTCCTCAATTTTCTTCAGATAATGCTAGATTAGAGATTAGAGATTCACACGAGGTAGAC 135
Db 35392 ACTTCCAAAGTATGTTTATAAAATATATGCTCTAAAACCAATTAAGTTGAGAAAATATATCTG 35451
Qy 136 ACCGAAAGAAACTTTGCCCCAATAATGCTTCTGTATTTTCATAGCAAGAGATTAAAGT 195
Db 35452 CATGAAAAAATTTTAAAGTAATTAATAGTAATCTGCAATGCTAATAATAACAGCTTTAAAT 35511
Qy 196 TTTCATTTAAGAGCCATTGTGAATTATACAACTAAATAAATAATGCAAGTGGATTTGAA 255
Db 35512 RGCTACCTGAGATGTCACTTGATGTTGATCAAAATATGAAAGGAGCTGGAGTTTAGTGAA 35571
Qy 256 CAGTCTCTTCTCTGAT 271
Db 35572 CACAGTCTCCAGTGT 35587

RESULT 11

US-10-741-601-5661/c
; Sequence 5661, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5661
; LENGTH: 101782
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(101782)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5661

Query Match 8.1%; Score 40.8; DB 20; Length 101782;
Best Local Similarity 50.5%; Pred. No. 27;
Matches 99; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 76 AATTCCTCAATTTTCTTCAGATAATGCTAGATTAGAGATTAGAGATTCACACGAGGTAGAC 135
Db 11322 ACTTCCAAAGTATGTTTATAAAATATATGCTCTAAAACCAATTAAGTTGAGAAAATATATCTG 11263
Qy 136 ACCGAAAGAAACTTTGCCCCAATAATGCTTCTGTATTTTCATAGCAAGAGATTAAAGT 195
Db 11262 CATGAAAAAATTTTAAAGTAATTAATAGTAATCTGCAATGCTAATAATAACAGCTTTAAAT 11203
Qy 196 TTTCATTTAAGAGCCATTGTGAATTATACAACTAAATAAATAATGCAAGTGGATTTGAA 255
Db 11202 RGCTACCTGAGATGTCACTTGATGCTGATCAAAATATGAAAGGAGCTGGAGTTTAGTGAA 11143
Qy 256 CAGTCTCTTCTCTGAT 271
Db 11142 CACAGTCTCCAGTGT 11127

RESULT 12

US-10-437-963-36653/c
; Sequence 36653, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 36653
; LENGTH: 1886
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_40458C.1
US-10-437-963-36653

```
Query Match      8.0%; Score 40.4; DB 20; Length 1886;
Best Local Similarity 53.9%; Pred. No. 6.2;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 114 AGAGATTCAACAGGGATAGACACGGAAGAAAACTTTGCCCAATAAGCTTTCTGGTAT 173
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 AAAAAATTGCACCAATGCTAGAAATCAGATGAGAGAGCTGTTCAATAAAGATATGCTGATAT 158
Qy 174 TTCATAGACAGAGATTTAAAGTTTCCATTTAAGAGCCATTTGAATTATACACACATA 233
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 ATACACATCAAGGGTTTCAAAATTTTCTACATTACTCCATTCATGATTTTCCCAATCACA 98
Qy 234 AAAAATGCAAGTGGATATTGAACAGTCTCTCTCTC 267
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 AGAACAGAAATCAGACATATGTTTCAGTTTGCTATC 64

RESULT 13
US-10-424-599-18984
; Sequence 18984, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 18984
; LENGTH: 797
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_117146C.1
US-10-424-599-18984

Query Match      8.0%; Score 40.2; DB 19; Length 797;
Best Local Similarity 48.5%; Pred. No. 4.9;
Matches 111; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 53 TTTAACTATGATACATATTGAAATCCCAATTTTCTTCAGATAATGTGATTAGAGAT 112
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 TTGAATTTTTTTTAAATAATTTATTAGAAATCTTTTATACATTTAATTTATTATTAAT 352
Qy 113 TAGAGATTCAACAGGGATAGACACCGAAGAAAACTTTGCCCAATAAGCTTTCTGGTA 172
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 TTGATAATTAAACAAATTTGATATTTAATTCGAATGTTGATTATAGATATTATTACAATA 412
Qy 173 TTTCATAGCAGAGATTTAAGTTTTCGATTTAAGAGCCATTTGTAATTATACACAAT 232
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 ATTAATAGTTAAAAAACTTTGGTATTTAATTTGAATAGTGTATTAGATGTTTATTACAGT 472
Qy 233 AAAAAATGCAAGTGGATATTGAACAGTCTCTCTCTCATTAATCTAAAT 281
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
473 GATTGAAAAATTAATAAATTTGGTATTAAATGTAATGATGATTATTAGAT 521

RESULT 14
US-10-473-126-240/c
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
```

```
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

Query Match      8.0%; Score 40; DB 21; Length 8056;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 109; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 15 TAGGAATATAACATTTTCAGTAGCATGCTGATACCACAGTTTAAACTATGCGATACATATTT 74
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7159 TAGCTTTACGATTTATATATATATATTTTATTATATTAAAAAAATTTATAATTTT 7100
Qy 75 GAATTCCAAATTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACAGGGATAGA 134
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7099 TAAATTAAAAAATTAATATTAAAAAATAAATAAATTTTAAATTTTAAATTTTAAATTTT 7040
Qy 135 CACGGAAGAAAACTTTGCCCAATAAGCTTTCTGGTATTTTCATAGCAGAGATTTTAAG 194
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7039 TTTTAAAAAATAAATACGAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTT 6980
Qy 195 TTTTCCATTTAAGAAGCCATTGTGAATTATACAAATAAAAAA 238
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6979 TTTTATATATATATCGTATTTTAAATAAATAAAAAAATAAAAAA 6936

RESULT 15
US-10-674-124A-1055/c
; Sequence 1055, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
; APPLICANT: INOKO, Hidetoshi
; APPLICANT: TAMURA, Gen
; TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
; TITLE OF INVENTION: GENETIC POLYMORPHISM MARKERS
; FILE REFERENCE: ORIN-003CIP
; CURRENT APPLICATION NUMBER: US/10/674,124A
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 10/257,511
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/JP00/07621
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: JP2000-112699
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: JP2002-327516
; PRIOR FILING DATE: 2002-09-28
; PRIOR APPLICATION NUMBER: JP2002-383869
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 1055
; LENGTH: 350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: chr1.fa.07fzr.118358138
; FEATURE:
; OTHER INFORMATION: Located on chromosome 1
; FEATURE:
; OTHER INFORMATION: Distance between a terminus base of telomere on
; OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
; OTHER INFORMATION: sequence : 111170513
; FEATURE:
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 158826
US-10-674-124A-1055

Query Match      7.9%; Score 39.8; DB 20; Length 350;
Best Local Similarity 50.8%; Pred. No. 4.3;
Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
```

Qy	51	CGTTTAAACTATGGGATACATATTTGAAATTCCAAATTTTCTTCAGATGAATGATAGAG	110
Db	208	CATGTAAACATATATACACATATTTTTTATATATATATATATACATATATTTTGTATATTT	149
Qy	111	ATTAGAGATTCAACACAGGGATAGACACCGAAGAAAATTTGCCCAAAATAAGCTTTTCGG	170
Db	148	ATTATATATATATATATATATATATATATATATATATTTCAACACAATTTCTTATTCGAG	89
Qy	171	TATTTTCATAAGCAAGAGATTTAAGTTTTCATTTTAAAGAGCCATTTGTGAATTTATACAACA	230
Db	88	TCATTTCACTATAGATACACTAGTGTGTTTTTATTTCTTGGCGTGTGTGAATAATGTCGCA	29
Qy	231	ATAAAAA	237
Db	28	ATGAATA	22

Search completed: November 8, 2005, 23:53:31
Job time : 652.857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 22:16:08 ; Search time 2598.77 Seconds
(without alignments)
7367.447 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320

Perfect score: 503

Sequence: 1 tagagtaagtaccttagaa.....atatttggaactgatgtct 503

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_hc.*
- 4: gb_est3.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_gss1.*
- 9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.6	10.3	1101	9	CNS00E47
C 2	48.2	9.6	1101	9	CNS0039G
C 3	46.4	9.2	747	8	BH957102
C 4	45.6	9.1	1106	9	CNS04LKI
C 5	44.6	8.9	766	8	BH460886
C 6	43.4	8.6	843	9	CNS00CS1
C 7	43.2	8.6	618	9	CL335225
C 8	43.2	8.6	618	9	CNS0167M
C 9	43.2	8.6	1201	9	CNS0167M
C 10	42.6	8.5	623	8	BH742342
C 11	42.4	8.4	669	8	BZ465736
C 12	42.4	8.4	818	8	BH661305
C 13	42.2	8.4	766	8	AQ544187
C 14	42.2	8.4	874	9	CNS019400
C 15	42.2	8.4	1101	9	CNS000B8
C 16	42	8.3	473	6	CNS000B8
C 17	42	8.3	959	9	CNS0170R
C 18	42	8.3	1101	9	CNS006X3
C 19	41.8	8.3	266	9	CL883428
C 20	41.8	8.3	496	8	BH711021
C 21	41.8	8.3	591	8	BH467264
C 22	41.4	8.2	448	7	CK732164
C 23	41.4	8.2	617	9	CL625152
C 24	41.4	8.2	628	9	CL611954

C 25	41.2	8.2	1174	9	CL110860
C 26	41	8.2	724	7	CV464338
C 27	41	8.2	841	9	CR812616
C 28	41	8.2	1086	9	CNS00YXK
C 29	41	8.2	1123	8	CC273164
C 30	40.8	8.1	321	4	BG438771
C 31	40.8	8.1	412	5	BP623936
C 32	40.8	8.1	538	5	BH811789
C 33	40.8	8.1	693	9	CL934409
C 34	40.8	8.1	777	6	CB180244
C 35	40.8	8.1	832	8	BZ601497
C 36	40.8	8.1	1101	9	CNS00EJ4
C 37	40.8	8.1	1101	9	CNS00EJ4
C 38	40.6	8.1	705	8	BH970382
C 39	40.6	8.1	734	9	CNS010MP
C 40	40.6	8.1	736	8	BZ469037
C 41	40.6	8.1	1101	9	CNS00Z6Z
C 42	40.4	8.0	515	2	BZ238266
C 43	40.4	8.0	738	6	CA104351
C 44	40.4	8.0	784	4	BI738064
C 45	40.2	8.0	326	4	BG370931

ALIGNMENTS

RESULT 1
LOCUS CNS00E47 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BAC28M18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL068720
VERSION AL068720.1 GI:4948863

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Rukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2, a cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BAC28M18"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 10.3%; Score 51.6; DB 9; Length 1101;
Best Local Similarity 29.8%; Pred. No. 0.011;


```

Qy 130 ATAGACACCGAAGAACTTTGGCCCAATTAAGCTTTCTGCTATTTTCATAGCAAGAGAT 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 ATAAAAATTTAACTAACATAAAGGCGACAAAAATAATTTTTTTTATAGTTGTACAT 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 190 TTAAGTTTTCAC 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TTAATTTTACAA 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
CNS041KI
LOCUS CNS041KI 1106 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
118P23 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL296235
VERSION AL296235.1 GI:9034815
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bertot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 1106)
GENOSCOPE Genoscope.
DIRECT SUBMISSION Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
location/Qualifiers
1..1106
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="118P23"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG118CH12LP1-end : T7"

ORIGIN
Query Match 9.1%; Score 45.6; DB 9; Length 1106;
Best Local Similarity 41.4%; Pred. No. 0.41;
Matches 116; Conservative 24; Mismatches 140; Indels 0; Gaps 0;

Qy 20 ATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACATGATGATACATTTGAATT 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 AAAAAAAATTTTCTTTTCTTTTAAACCAACTTACACACAAATTTTWTWAAAA 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 80 CCAATTTTCTTCAGTAATGTGATTAGAGATTAGAGATTCAACCGGATAGACCG 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 837 ACAAAWAATTTTTCACAAATTTWCAAAAAAAATTTWTATCATCAAAAAACAACCT 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 140 AAAGAAAATTTGGCCCAATTAAGCTTTCTGCTATTTTCATAGCAAGAGATTTAAGTTTC 199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 YTTAAAACTTTTAAAAAHYTTTTTATWACCTTTTAACTTTTAAAAAACYYTTT 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 200 CATTTAAGAGCCATTTGTGAATTATACAACAATAAAAAATGCAAGTGATATTGAACAGT 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 TAAAAACACCACTCTCCYACAAACWCAAAAAATAAAAAAATTAATTAATTAATTTT 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 260 CTCTTCTCTGATAATTTTAAATACAGTACAGTTCACGCC 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 TTTBTTTTTTTTTTATTTTATTTAAAWAAAAACAACACCC 1056
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
BH460886/c
LOCUS BH460886 766 bp DNA linear GSS 13-DEC-2001
DEFINITION BOGSK22TF BOGS Brassica oleracea genomic clone BOGSK22, genomic
survey sequence.
ACCESSION BH460886
VERSION BH460886.1 GI:17650631
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 766)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOGSK22TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
source
location/Qualifiers
1..766
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGSK22"
/clone_lib="BOGS"
/note="Vector: pHS1, Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS1 using BstXI linkers"

ORIGIN
Query Match 8.9%; Score 44.6; DB 8; Length 766;
Best Local Similarity 49.4%; Pred. No. 0.7;
Matches 116; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 27 ATTTTCAGTAGCATGCTGATACCAACGTTTAAACTATGATACATATTGAAATTCCAAAT 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 708 ATAACTATATATTTTGTATATAATTTTATTAAGTGTTTATTTCAAAATGTAATTAATA 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 87 TTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCGGATAGACACCGAAGAAA 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 648 ATATATAACATATTTCTGTTAGATATAACATATGTAACAACTTGTAAAAATTAATAAAA 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 147 ACTTTGCCCAATTAAGCTTTCTGCTATTTTCATAGCAAGAGATTTAAGTTTCCATTTAA 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 CAATGCCGAAATATTTTGTCTCAAAATTTTCAATTTTGTATTATATATACATATTT 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 207 GAAGCCATTGTGAATTTACAAACAATAAAAAATGCAAGTGGATATTGAACAGTCT 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 TAGGAATAATATACATAAACTATGAGGACCAACTAATTTGATTATTGAACTTTT 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
LOCUS
CNS00CS1
DEFINITION
  CNS00CS1
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
  843 bp DNA linear GSS 04-JUN-1999
ACCESSION
  AL059666.1 GI:4947129
VERSION
  AL059666
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 843)
  Direct Submission
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazuo Osoegawa and
  Aaron Mamonos in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain y2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
  source
    1..843
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"
    /clone="BACR26H19"
    /clone_lib="RPCI-98"
    /note="end : TET3"
ORIGIN
  Query Match 8.6%; Score 43.4; DB 9; Length 843;
  Best Local Similarity 36.7%; Pred. No. 1.5;
  Matches 90; Conservative 39; Mismatches 116; Indels 0; Gaps 0;
  QY 38 ATGCTGATACCAAGTTAACTAATGATGATACATATTTGAATTCCTCAATTTTCTTCGAT 97
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  578 WTTTATATATWATTTTTTTTTTTTTTTTTTAYAWAAATTAATTTATATATTTATATAT 637
  QY 98 AATGCTGATAGATTAGAGATTCACACGCGATAGACACGGAAGAAACTTTGCCCA 157
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  638 WTTATWAAWAAWATTAATWAAWATAAATAAATAWAAWAAWAAWAAWAAWAAWAAWAAW 697
  QY 158 ATAAGCTTTTCGGTATTCATAAGCAGAGATTTAAGTTTTCATTTTAAGACCCATTGT 217
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  698 TATATATATWATTAATWAAWATTAATWAAWATAWATAWATAWATAWATAWATAWATAW 757
  QY 218 GAATTAACACATATAAATGCAAGTGGATATTGAACAGTCTCTCTCTGATATTTCT 277
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  758 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 817
  QY 278 AATA 282
  Db : : : : :
  818 ATATW 822

```

RESULT 7

```

CC764640/c
LOCUS
DEFINITION
  CC764640
  CH240_47L12.TJ CHORI-240 Bos taurus genomic clone CH240_47L12,
  genomic survey sequence.
  495 bp DNA linear GSS 27-JUN-2003
ACCESSION
  CC764640
VERSION
  CC764640.1 GI:32311143
KEYWORDS
  GSS.
SOURCE
  Bos taurus (cow)
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovinae; Bos.
REFERENCE
  1 (bases 1 to 495)
  Larkin, D.M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P.,
  Bachman, S., Green, S., Campos, B.J., Benson, L.D., Edwards, J., Liu, L.,
  Womack, J.E., de Jong, P.J. and Lewin, H.A.
  Bovine BAC end sequences from CHORI-240 library
  Unpublished (2003)
  Other GSSs: CH240_47L12.TV
  Contact: Harris Lewin
  Department of Animal Sciences
  University of Illinois at Urbana Champaign
  1201 W. Gregory Dr., Urbana, IL 61801, USA
  Tel: 217 333 5998
  Fax: 217 244 5617
  Email: h-lewin@uiuc.edu
  Clones are derived from the bovine BAC library CHORI-240
  (http://www.chori.org/bacpac/bovine240.htm). For BAC library
  availability, please contact Pieter de Jong (pdejong@email.cho.org).
  Clones may be purchased from BACPAC Resources
  (http://www.chori.org/bacpac/ordering/information.htm). This work
  was undertaken as part of the International Bovine BAC Mapping
  Consortium (IBBMC) by University of Illinois at Urbana
  Champaign, USA with funds provided by grant NO. AG202-34480-11828
  from USDA-CSRES and AG99-35205-8534 from USDA/NRI (Livestock
  Genome Sequencing Initiative)
  Plate: 47 row: L column: 12
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
    1..495
    /organism="Bos taurus"
    /mol_type="genomic DNA"
    /strain="breed: Hereford"
    /db_xref="taxon:9913"
    /clone="CH240_47L12"
    /sex="Male"
    /cell_type="Blood"
    /clone_lib="CHORI-240"
    /notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
    Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
    library (Male) produced by Pieter de Jong"
ORIGIN
  Query Match 8.6%; Score 43.2; DB 9; Length 495;
  Best Local Similarity 50.0%; Pred. No. 1.5;
  Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
  QY 21 TATAACATTTCTAGTACGCTGATACCAACGTTTAACTATGATACATATTTGAATTC 80
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  452 TTTAAATTTTCAAAAAATGCACAAAATATGTAATAATTTTGAACATCTTTAAATT 393
  QY 81 CAAATTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACACGGGATAGACACCGA 140
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  392 CAAAGTAACATCTGTGTAATGTGGTAATGCTGCTGGAGTCATTTAGGTATATCAACAGA 333
  QY 141 AAGAAATCTTTGCCAAATAAGCTTTCTGCTATTTTCATAGCAAGAGATTTAAGTTTTC 200
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  332 GAGAAGATATATATAAATATGCTTTTATATGATTGAGCTCTTATAGGTTCTATCTCTAT 273
  QY 201 ATTGAAGAGCCATTTGTGAATTATATACAAATAAAA 236
  Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  272 TTTCAAGTATTTCAATTTAATAGTGTATAAATTTAA 237

```

```

RESULT 8
CL335225
LOCUS
DEFINITION
  CL335225
  RPCI44_253A15, r RPCI-44 Sus scrofa genomic clone RPCI44_253A15,
  genomic survey sequence.
ACCESSION
  CL335225
VERSION
  CL335225.1 GI:51387193
KEYWORDS
  GSS.
SOURCE
  Sus scrofa (pig)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
  1 (bases 1 to 618)
  Rogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,
  Beever, J.E. and Schook, L.B.
  Piggy-BACing the Human Genome: Constructing a Porcine Physical Map
  Through Comparative Genomics
  Unpublished (2004)
  Other GSSs: RPCI44_253A15.f
  Contact: Lawrence B. Schook
  Department of Animal Sciences
  University of Illinois at Urbana Champaign
  1201 W. Gregory Dr., Urbana, IL 61801, USA
  Tel: 217 265 5326
  Fax: 217 244 5617
  Email: schook@uiuc.edu
  Clones are derived from the porcine BAC library RPCI-44
  (http://www.bacpac.chori.org/porcine242.htm). For BAC library
  availability, please contact Pieter de Jong (pdejong@chori.org).
  Clones may be purchased from BACPAC Resources
  (http://BACPACresources.chori.org). This work was undertaken as part
  of the International Swine Genome Sequencing Consortium by
  University of Illinois at Urbana Champaign, USA with funds provided
  by grant No. AG2002-34480-11828 from USDA-CSREES and
  AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
  Initiative)
  Plate: 253
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
      1..618
        /organism="Sus scrofa"
        /mol_type="genomic DNA"
        /strain="four pigs (breed: 37.5% Yorks Landrace and 25%
        Meishan)"
        /db_xref="taxon:9823"
        /clones="RPCI44_253A15"
        /sex="male"
        /cell_type="blood"
        /clone_lib="RPCI-44"
        /note="Vector: pTARBAC2; Site 1: EcoRI; Site 2: EcoRI;
        porcine male BAC library produced by Pieter de Jong"
ORIGIN
  Query Match      8.6%; Score 43.2; DB 9; Length 618;
  Best Local Similarity 48.3%; Pred. No. 1.6;
  Matches 117; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy 53 TTTAACTAGGATACATATTTGAAATTCCTCAATTTTCTTCAGATATGATAGAT 112
Db 357 TTTATTTTCAATTTTACCAATCTCTCTCAAAATAATACAGAGGGAATTTTATTTAATA 416
Qy 113 TAGAGATTCACACAGGATACACACCGAAGAACTTTGCCAAATAAGCTTCTCGTA 172
Db 417 TGTAAATTTAGCTAAGTATAGTTAATTAAGTCAACACAGCACTGGCCATGAGTTA 476
Qy 173 TTTCATAGCAAGAGATTTTAAAGTTTTCATTTTGAAGAGCCATTGTGAATTATCAACAAT 232
Db 477 CTTATATAATTCAGAAAAAGAGCTTACTTATTTGCAATTATTACTTTTCTATCATACAGCAGA 536
Qy 233 AAAAAATGCAGTGGATATTGAACAGCTCTCTTCTGTGATTAATTTCTAAATACAGTACAGTT 292

```

```

Db 537 AAACAATGTACATATTATTTCTCAATTTGCAATATGAAAAGTCTATTTCATAANNNGTT 596
Qy 293 CA 294
Db 597 CA 598

RESULT 9
CNS0167M/c
LOCUS
DEFINITION
  CNS0167M
  Drosophila melanogaster genome survey sequence T7 end of BAC
  BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL106396
VERSION
  AL106396.1 GI:5621701
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
ORGANISM
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1201)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelOBAC11.
FEATURES
  source
    Location/Qualifiers
      1..1201
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACN15M24"
        /clone_lib="DrosBAC"
        /plasmid="pBelOBAC11"
        /note="end : T7"
ORIGIN
  Query Match      8.6%; Score 43.2; DB 9; Length 1201;
  Best Local Similarity 41.5%; Pred. No. 1.8;
  Matches 102; Conservative 22; Mismatches 122; Indels 0; Gaps 0;

Qy 44 ATACCAAGTTTAAACATCGATACATATTTGAATTCCAAATTTTCTTCAGATAATGTG 103
Db 949 AAAAAAATATTAATAATTTTATTAATTAATAATAATTTTATTAATAATAATAATAA 890
Qy 104 ATTAGAGATTAGAGATTCAACACGAGGATAGACACCGAAGAAACTTTGCCCAATATAGC 163
Db 889 TTAATAATTTAATAATTTAATAATAATAATAATAATAATAATAATAATAATAATAAT 830
Qy 164 TTTCTGTGATTTCAATAGCAAGAGATTTAAGTTTTCATTTTAAGACCATTTGGAATTA 223
Db 829 TATAAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 770
Qy 224 TACAACAATAATAATAATGCAAGTGGATATTCACACAGTCTCTCTCTGTGATTTCTTAAATAC 283
Db 769 AAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 710
Qy 284 AGTACA 289
Db 709 AAAAAA 704

RESULT 10

```

```

BH742342      BH742342      623 bp      DNA      linear      GSS 25-FEB-2002
LOCUS          gu74b05.g1 BoBuds01 Brassica oleracea genomic clone gu74b05 5',
DEFINITION     genomic survey sequence.
ACCESSION      BH742342
VERSION        BH742342.1 GI:18876955
KEYWORDS       GSS.
SOURCE         Brassica oleracea
ORGANISM       Brassica oleracea
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
                1 (bases 1 to 623)
                Katari,M., O'Shaughnessy,A., Palmer,L., Bahret,A., Baker,J.,
                Balija,V., Cunniss,D.M., Katzenberger,F., King,L., Kirchoff,K.,
                Kuit,K., Miller,B., Muller,S., Nascimento,L., Preston,R.,
                Santos,L., Shah,R., Zutavern,T., Dedhia,N., Rabinowicz,P.D. and
                McCombie,W.R.
                Whole Genome Shotgun Reads from Brassica oleracea (2002b)
TITLE          Unpublished (2002)
JOURNAL        Contact: W. Richard McCombie
COMMENT        Lita Annenberg Hazen Genome Sequencing Center
                Cold Spring Harbor Laboratory
                PO Box 100, Cold Spring Harbor, NY 11724, USA
                Tel: 516 367 8884
                Fax: 516 367 8874
                Email: mcombie@cshl.org
                Plate: gu74 row: b column: 05
                Seq primer: -2lUnivRev
                Class: shotgun
                High quality sequence stop: 623.
FEATURES       source
                Location/Qualifiers
                1..623
                /organism="Brassica oleracea"
                /mol_type="genomic DNA"
                /db_xref="taxon:3712"
                /clones="gu74b05"
                /clone_lib="BoBuds01"
                /note="Vector: M13 for .x reads, pBluescript for .b and .g
                reads; Site1: EcoRV; Whole genome shotgun library from
                flowering buds. DNA was purified from a crude nuclear prep
                using Brassica oleracea TO1000DH3 buds provided by Thomas
                Osborn at the University of Wisconsin. Genomic DNA
                provided by Pablo Rabinowicz (CSHL) and shotgun library
                prepared in McCombie Lab."
ORIGIN
Query Match      8.5%; Score 42.6; DB 8; Length 623;
Best Local Similarity 47.5%; Pred. No. 2.3;
Matches 126; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

Qy 11 ACCTTAGGATATTAACATTTTCAGTAGCATGCTGATACCAAGCTTTAAACTATGATACAT 70
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 ACCTTATAAATGACAAATTTCCATATAAAATTTATTACGAAATTCGCTATATTTAGCTACCT 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 71 ATTGGAATCCAAATTTTCTTCAGATAAATGTGATTAGAGATTAGAGATTCAACACGGGA 130
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 AATAAATTTAGAGATATGCATCCTTCATTTAATTAATCAAAATCAAAATTCGATT 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 131 TAGACACGGAAGAAAACCTTTGCCCAATAAGCTTTCTGGTATTTTCAATAAGCAAGAGATT 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 TTGCAATCCATAAAACTATAAAATACATTAATAATTCAGATTAAATTCATTGTTGTTATT 371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 191 TAAGTTTTCATTTAAGAACGCATTTGAATTATACAAATATAAAATCAAGTGGATA 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 GAAATTAATAATCAATCAAAATCATTTAAATTTTACAAATATAAAATAAATACTCAGTTATA 431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 251 TTGAACAGTCTCTCTCTGATAATT 275
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 TTTAATTAATAGTCTCTCTTAATT 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11

```

```

BZ465736      BZ465736      669 bp      DNA      linear      GSS 13-DEC-2002
LOCUS          BONMQ35TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONMQ35,
DEFINITION     genomic survey sequence.
ACCESSION      BZ465736
VERSION        BZ465736.1 GI:26757807
KEYWORDS       GSS.
SOURCE         Brassica oleracea
ORGANISM       Brassica oleracea
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
                1 (bases 1 to 669)
                Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
                Whole genome shotgun sequencing of Brassica oleracea
                Unpublished (2001)
JOURNAL        Other_GSSs: BONMQ35TR
COMMENT        Contact: Chris Town
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TF
                Class: sheared ends.
FEATURES       source
                Location/Qualifiers
                1..669
                /organism="Brassica oleracea"
                /mol_type="genomic DNA"
                /strain="TO1000DH3"
                /db_xref="taxon:3712"
                /clone="BONMQ35"
                /clone_lib="BO_1.6_2_KB_tot"
                /note="Vector: pHOsi; Site 1: BstXI; 1.6-2 kb sheared
                total DNA inserted into pHOsi using BstXI linkers"
ORIGIN
Query Match      8.4%; Score 42.4; DB 8; Length 669;
Best Local Similarity 50.5%; Pred. No. 2.6;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 6  TAAGTACCTTAGGAATATAACATTTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGA 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 451 TAGTACCTTAAATATATATTTTCTTTAAATAATCATATAAAACAAAATTTTACACTTTATA 510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 66 TACATATTTGAATTCCAAATTTTCTTCAGATAAATGTGATTAGAGATTAGAGATTCACCC 125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 511 TACATATTTCAAATCAAAATAAATAATTTAAATTTGATTTATATCAAAAATTTGATTCAAA 570
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 126 AGGGATAGACACCCGAAAGAAACTTTGCCCAATAAGCTTTCTGGTATTTTCATAAGCAAG 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 ATATACATATGTTTAAATAATAAATTTTACTAAATAATTTTCCAAATAACCATTTATAA 630
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 186 AGATTTTAAGTTTTCATTTTAAGAA 209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 AAATGGTTTCAATATATATAAGAA 654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
BH661305/c    BH661305/c    818 bp      DNA      linear      GSS 19-FEB-2002
LOCUS          BOHYL59TF BO_2.3_KB Brassica oleracea genomic clone BOHYL59,
DEFINITION     genomic survey sequence.
ACCESSION      BH661305
VERSION        BH661305.1 GI:18720067
KEYWORDS       GSS.
SOURCE         Brassica oleracea
ORGANISM       Brassica oleracea
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
                1 (bases 1 to 818)

```

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of *Brassica oleracea*
JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.

FEATURES
 source
 Location/Qualifiers
 1. .818
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOHLS9"
 /clone_lib="BO 2.3 KB"
 /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 8.4%; Score 42.4; DB 8; Length 818;
 Best Local Similarity 47.7%; Pred. No. 2.7;
 Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 28 TTTCAGTACGATGCTGATACCAACGTTTAACTATGATGATACATATTTGCAATTTCCAAATTT 87
 DB 301 TTTCAACTAACCAATAAACAATTTAACTTTATATATACATTTTCAATCCAAATATAA 242

QY 88 TTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACGAGGATAGACACCGAAGAAAA 147
 DB 241 TAAATCAAGTTGATTATATACAAAAATTGATTCAAAAATATATATATATATATATAT 182

QY 148 CTTTGCCCAATAAGCTTTCTGGTATTTCAATGCAAGAGATTTAAGTTTCCATTTAAG 207
 DB 181 GGATTTCTACTAAACTATTTTTCATAACCACTTAAAAAAATGTTATCAATATATATAAG 122

QY 208 AAGCCATGTGATTTATACAAATAAATAAATGCAAGTGGATTTGAACAGCTCTCTCTC 267
 DB 121 AAAATATATACAGAGCCCAATTTGAATACCACTCAAAATATGTTTATATATTTTA 62

QY 268 TGATAATTTCTAAATACAGTA 287
 DB 61 TATTAAGTTTATAAATATA 42

RESULT 13
LOCUS AQ544187/c
DEFINITION RPCI-11-326P5, TJ RPCI-11 Homo sapiens genomic clone RPCI-11-326P5, genomic survey sequence.
ACCESSION AQ544187
VERSION AQ544187.1 GI:4968489
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1. (bases 1 to 766)
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P., and Venter, J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: RPCI-11-326P5.TV
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200

Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1. .766
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GDB:7625164"
 /db_xref="taxon:9606"
 /clone="RPCI-11-326P5"
 /sex="Male"
 /cell_type="Lymphocytes"
 /clone_lib="RPCI-11"
 /note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 8.4%; Score 42.2; DB 8; Length 766;
 Best Local Similarity 49.8%; Pred. No. 3;
 Matches 107; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 59 CTATGGATACATATTTGCAATTTCCAAATTTTCTTTCAGATAATGTGATGAGATTAGAGA 118
 DB 526 CTATTGCTGTATAAACAATTTATTTTCTTCTACTATAATATACCTAGAAATATATA 467

QY 119 TTCAACGAGGATAGACACCGAAGAAACATTTTCCCAAAATAGCTTTCTCGTATTTTCAT 178
 DB 466 TTTATACAAACATTTATTCATTGTGAAAAATGTAATCTTTTAAATAGTCTCTGAGATTAA 407

QY 179 AAGCAAGAGATTTAAGTTTTCATTTAAGAGCCATTTGTAATTATACAAATAAAAAA 238
 DB 406 AATTAGCAGATCAAGGCTATAAATTTTATAGGCTCTTGATTTATATATTGATAAAATTG 347

QY 239 TGCAAGTGGATATTGAACAGCTCTCTCTCGATAA 273
 DB 346 CTCCACAAATATATAACAATTTAGTGTCACATTA 312

RESULT 14
LOCUS CG019400
DEFINITION ZUAAC48TV ZM_3.0_4.0 KB Zea mays genomic clone ZMMBP00005G23, genomic survey sequence.
ACCESSION CG019400
VERSION CG019400.1 GI:33891565
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
AUTHORS 1. (bases 1 to 874)
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Sudman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Numberg, A., Robbins, D. and Lakey, N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other GSSs: ZUAAC48TH
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TP
 Class: sheared ends.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:46:48 ; Search time 101.246 Seconds
(without alignments)
7308.644 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125

Perfect score: 125

Sequence: 1 cccctgccctgccctgcc.....gcggggcagcggggcccagc 125

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

1: Geneseq19806.*

2: Geneseq19908.*

3: Geneseq20008.*

4: Geneseq2001as.*

5: Geneseq2001bs.*

6: Geneseq2002as.*

7: Geneseq2002bs.*

8: Geneseq2003as.*

9: Geneseq2003bs.*

10: Geneseq2003cs.*

11: Geneseq2003ds.*

12: Geneseq2004as.*

13: Geneseq2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125	100.0	503	2	Aax19032 Human PPA
2	118.6	94.9	12548	13	Adg36487 Human aut
3	118.6	94.9	158417	13	Adg36461 Human aut
4	81.4	65.1	201	2	Aax19063 Human PPA
5	57.8	46.2	1416	8	Abz20967 Animal te
6	56.6	45.3	420	13	Acn51217 Cotton an
7	54.4	43.5	1315	2	Aaz41132 Human nor
8	54.2	43.4	200	5	Abv61578 Human pro
9	54.2	43.4	629	13	Acn54594 Cotton an
10	53	42.4	1131	12	Adq23807 Human sof
11	52.2	41.8	168	6	Abn69641 Streptoco
12	52	41.6	3163	10	Adc87060 Human GPC
13	52	41.6	28198	10	Adg37080 Mouse pla
14	51.8	41.4	344	4	AI182007 Human pol
15	51.8	41.4	437	13	Acn58918 Cotton gy
16	51.4	41.1	110000	12	Adq97050 Human can
17	51.2	41.0	349	4	AI184504 Human pol
18	51	40.8	588	13	Acn54596 Cotton an
19	51	40.8	600	6	Abq52497 Oligonuci
20	51	40.8	600	6	Abq52496 Oligonuci

21	51	40.8	2188	2	AAZ77506	Human ova
22	50.8	40.6	142	7	ADR41362	Human CD-
23	50.8	40.6	1045	6	ABI199807	Mouse isc
24	50.8	40.6	1286	6	ABI199656	Mouse isc
25	50.6	40.5	434	4	AAI84343	Human pol
26	50.6	40.5	3198	2	AAx02974	Human il-
27	50.4	40.3	594	6	ABQ43958	Oligonuci
28	50.4	40.3	594	6	ABQ43959	Oligonuci
29	50.2	40.2	1845	3	AAZ98400	Canine be
30	50.2	40.2	1845	6	ABK40732	Dog betai
31	50.2	40.2	43058	6	ABL64982	Lung canc
32	50.2	40.2	43058	6	ABL65219	Lung canc
33	50.2	40.2	43058	6	ABN97455	Gene #395
34	50	40.0	615	8	ACA23975	Prokaryot
35	50	40.0	1948	6	ABQ69181	Listeria
36	50	40.0	5082	6	ABQ70953	Listeria
37	49.8	39.8	204	13	ADR93549	Novel S.
38	49.8	39.8	209	13	ADR91550	Novel S.
39	49.8	39.8	282	13	ADR91549	Novel S.
40	49.8	39.8	308	13	ADR93088	Novel S.
41	49.8	39.8	612	13	ADR93335	Novel S.
42	49.8	39.8	884	8	ACC48196	EMCV inte
43	49.8	39.8	1387	10	ADC87462	Human GPC
44	49.8	39.8	2685	6	ABK83742	Human cDN
45	49.8	39.8	12733	6	ABK98631	Vector pE

ALIGNMENTS

RESULT 1

AAx19032

ID AAX19032 standard; DNA; 503 BP.

XX AAX19032;

AC AAX19032;

DT 13-MAY-1999 (first entry)

XX Human PPAR-gamma-1 proximal promoter, exon A1 and intron A1.

XX Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;

KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;

KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;

KW insulin resistance; non-insulin-dependent diabetes mellitus;

KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;

KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.

OS Homo sapiens.

XX WO9905161-A1.

XX 04-FEB-1999.

XX 24-JUL-1998; 98WO-US015411.

XX 25-JUL-1997; 97US-0053692P.

XX (LIGA-) LIGAND PHARM INC.

XX (INSP) INST PASTEUR.

XX Briggs MR, Saladin RS, Auwerx J, Pajas L;

XX WPI; 1999-142844/12.

XX Newly isolated nucleic acid comprising a control region of a human

XX peroxisome proliferator activated receptor (PPAR) gamma gene - useful for

XX identifying modulators that are useful in treating diseases associated

XX with abnormal levels of human PPAR-gamma gene expression.

XX Claim 9; Page 78; 102pp; English.

XX The present invention describes an isolated, purified or enriched nucleic

XX acid comprising a control region of a human peroxisome proliferator

activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful for screening for agents capable of modulating the expression of a human PPAR-gamma gene. These agents (modulators) form pharmaceutical compositions that are useful for treating diseases associated with high/low levels of human PPAR-gamma gene expression. The diseases include obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas, abnormalities associated with anti-human immunodeficiency virus (HIV) treatment, insulin resistance, non-insulin-dependent diabetes mellitus (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI) tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis and bowel cancer. The nucleic acids are useful for studying the role of the PPAR-gamma gene in various diseases and disorders. The structure of PPAR-gamma enables genetic studies of PPAR-gamma mutations in humans, and evaluation of its role in disorders like insulin resistance, NIDDM, and diseases associated with altered adipose tissue function, like obesity and lipodystrophic syndromes. The nucleic acids are also useful for gene therapy and the production of transgenic animals, which are useful in screening assays. The control regions of the nucleic acids enable screening for modulators of the human PPAR-gamma gene, which are useful in designing drugs for treating disorders or diseases associated with the level of PPAR-gamma gene expression. The present sequence represents human PPAR-gamma-1 proximal promoter, exon A1 and intron A1

XX
SQ Sequence 503 BP; 60 A; 184 C; 211 G; 48 T; 0 U; 0 Other;

Query Match 100.0%; Score 125; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCTGCGCCCTGCCCCCAGCCACCCACCCACCCACCCAGCGGCGCGGC 60
Db 1 CCCCCTGCGCCCTGCCCCCAGCCACCCACCCACCCACCCAGCGGCGCGGC 60

Qy 61 GCCCGCCCCCGCGGCGCGGCTCGGCCGACCCCGGTTCCGCGGCGGCGGCG 120
Db 61 GCCCGCCCCCGCGGCGCGGCTCGGCCGACCCCGGTTCCGCGGCGGCGGCG 120

Qy 121 CCAGC 125
Db 121 CCAGC 125

RESULT 2
ADS36487/C
ID ADS36487 standard; DNA; 12548 BP.
XX
AC ADS36487;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human autoimmune disease-related genomic DNA sequence - SEQ ID 1701.
XX
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
XX
OS Homo sapiens.
XX
FN WO2004083403-A2.
XX
PD 30-SEP-2004.
XX
PF 18-MAR-2004; 2004WO-US008461.
XX
PR 18-MAR-2003; 2003US-0455444P.
XX
PR 25-APR-2003; 2003US-0465241P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cargill M, Begovich AB, Alexander HC;

XX
DR WPI; 2004-728480/71.
XX
PT New isolated nucleic acid molecule comprises at least 8 contiguous nucleotides where one of the nucleotides is a single nucleotide polymorphism (SNP), useful for diagnosing or treating autoimmune diseases, e.g. rheumatoid arthritis.
XX
PS Claim 16; SEQ ID NO 1701; 123pp; English.
XX
CC The invention comprises amino acid and coding sequences containing genetic polymorphisms associated with an altered risk of developing an autoimmune disease (e.g. rheumatoid arthritis). The invention further comprises a method of identifying an individual that has an altered risk of developing an autoimmune disease, comprising detecting a single nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA and protein sequences of the invention are useful for diagnosing and treating autoimmune diseases, such as: rheumatoid arthritis, type 1 diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease, myocarditis, Sjogren's disease, or primary systemic vasculitis. The present nucleic acid represents a human autoimmune disease-related genomic DNA sequence of the invention. NOTE: The present sequence is not shown in the specification, but has been retrieved from the WIPO website.

XX
SQ Sequence 12548 BP; 3528 A; 2769 C; 2492 G; 3757 T; 0 U; 2 Other;

Query Match 94.9%; Score 118.6; DB 13; Length 12548;
Best Local Similarity 96.8%; Pred. No. 2.8e-10;
Matches 121; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCTGCGCCCTGCCCCCAGCCACCCACCCACCCACCCAGCGGCGCGGC 60
Db 7139 CCCCCCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7080

Qy 61 GCCCGCCCCCGCGGCGCGGCTCGGCCGACCCCGGTTCCGCGGCGGCGGCG 120
Db 7079 GCCCGCCCCCGCGGCGCGGCTCGGCCGACCCCGGTTCCGCGGCGGCGG 7020

Qy 121 CCAGC 125
Db 7019 CCAGC 7015

RESULT 3
ADS36461
ID ADS36461 standard; DNA; 158417 BP.
XX
AC ADS36461;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human autoimmune disease-related genomic DNA sequence - SEQ ID 1675.
XX
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
XX
OS Homo sapiens.
XX
FN WO2004083403-A2.
XX
PD 30-SEP-2004.
XX
PF 18-MAR-2004; 2004WO-US008461.
XX
PR 18-MAR-2003; 2003US-0455444P.
XX
PR 25-APR-2003; 2003US-0465241P.
XX

PA	(APPL-) APPLERA CORP.
XX	
PI	Cargill M, Begovich AB, Alexander HC;
XX	
DR	WPI; 2004-728480/71.
XX	
XX	New isolated nucleic acid molecule comprises at least 8 contiguous
PT	nucleotides where one of the nucleotides is a single nucleotide
PT	polymorphism (SNP), useful for diagnosing or treating autoimmune
PT	diseases, e.g. rheumatoid arthritis.
XX	
PS	Claim 16; SEQ ID NO 1675; 123pp; English.
XX	
CC	The invention comprises amino acid and coding sequences containing
CC	genetic polymorphisms associated with an altered risk of developing an
CC	autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC	comprises a method of identifying an individual that has an altered risk
CC	of developing an autoimmune disease, comprising detecting a single
CC	nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC	and protein sequences of the invention are useful for diagnosing and
CC	treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC	diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC	bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC	anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease, Crohn's
CC	myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC	present nucleic acid represents a human autoimmune disease-related
CC	genomic DNA sequence of the invention. NOTE: The present sequence is not
CC	shown in the specification, but has been retrieved from the WIPO website.
XX	
SQ	Sequence 158417 BP; 47887 A; 30186 C; 31475 G; 48577 T; 0 U; 292 Other;
	Query Match 94.9%; Score 118.6; DB 13; Length 158417;
	Best Local Similarity 96.8%; Pred. No. 1.4e-10;
	Matches 121; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	1 CCCCTGCCCCCTGGCCTGCCACCCCACCCAGCCACCCACCCACCCAGCGGCCTCCGC 60
DB	5890 CCCTCCGCCCCGCCCCGCCCCCACCCACCCACCCACCCACCCAGCGGCCTCCGC 5949
QY	61 GCCCGCCCCGCGCGCGCGCGCTCGCGCCGACCCGCTTCGCGCGCGGAGCGGGGC 120
DB	5950 GCCCGCCCCGCGCGCGCGCGCTCGCGCCGACCCGCTTCGCGCGCGGAGCGGGGC 6009
QY	121 CCAGC 125
DB	6010 CCAGC 6014
	RESULT 4
ID	AAX19063
XX	AAX19063 standard; DNA; 201 BP.
XX	
AC	AAX19063;
XX	
DT	13-MAY-1999 (first entry)
XX	
DE	Human PPAR-gamma-1 proximal promoter.
XX	
KW	Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
XN	regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW	lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW	insulin resistance; non-insulin-dependent diabetes mellitus;
KW	polycystic ovary syndrome; Crohn's disease; gastrointestinal tract; Crohn's disease;
KW	inflammatory bowel disease; ulcerative colitis; bowel cancer; 89.
OS	Homo sapiens.
XX	
PN	WO9905161-A1.
XX	
XX	04-FEB-1999.
PD	
XX	
PF	24-JUL-1998; 98WO-US015411.
XX	

PR	25-JUL-1997; 97US-0053692P.
XX	
PA	(LIGA-) LIGAND PHARM INC.
PA	(INSP) INST PASTEUR.
XX	
PI	Briggs MR, Saladin RS, Auwerx J, Fajas L;
XX	
DR	WPI; 1999-142844/12.
XX	
PT	Newly isolated nucleic acid comprising a control region of a human
PT	peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT	identifying modulators that are useful in treating diseases associated
PT	with abnormal levels of human PPAR-gamma gene expression.
XX	
PS	Disclosure; Page 87; 102pp; English.
XX	
CC	The present invention describes an isolated, purified or enriched nucleic
CC	acid comprising a control region of a human peroxisome proliferator
CC	activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC	for screening for agents capable of modulating the expression of a human
CC	PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC	compositions that are useful for treating diseases associated with
CC	high/low levels of human PPAR-gamma gene expression. The diseases include
CC	obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC	abnormalities associated with anti-human immunodeficiency virus (HIV)
CC	treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC	(NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC	tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC	and bowel cancer. The nucleic acids are useful for studying the role of
CC	the PPAR-gamma gene in various diseases and disorders. The structure of
CC	PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC	and evaluation of its role in disorders like insulin resistance, NIDDM,
CC	and diseases associated with altered adipose tissue function, like
CC	obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC	for gene therapy and the production of transgenic animals, which are
CC	useful in screening assays. The control regions of the nucleic acids
CC	enable screening for modulators of the human PPAR-gamma gene, which are
CC	useful in designing drugs for treating disorders or diseases associated
CC	with the level of PPAR-gamma gene expression. The present sequence
CC	represents the human PPAR-gamma-1 proximal promoter
XX	
SQ	Sequence 201 BP; 23 A; 95 C; 73 G; 10 T; 0 U; 0 Other;
	Query Match 65.1%; Score 81.4; DB 2; Length 201;
	Best Local Similarity 96.0%; Pred. No. 0.00032;
	Matches 95; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY	30 ACCCCACACCCCAACCCCAAGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 69
DB	1 ACCCCACACCCCAACCCCAAGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 60
QY	90 CCACACCCGGTT---CCGCCCGCGGAGCGGGGCCCGC 125
DB	61 CCACACCCGGATCCGCCCGCGGAGCGGGGCCCGC 99
	RESULT 5
ID	ABZ20967/C
XX	ABZ20967 standard; DNA; 1416 BP.
XX	
AC	ABZ20967;
XX	
DT	29-APR-2003 (first entry)
XX	
DE	Animal test kit marker DNA 127-sp6.
XX	
KW	DNA marker; meat; consumer protection; genetic fingerprint; SNP;
KW	single nucleotide polymorphism; animal breeding; gene; ds.
OS	Bos taurus.
XX	
FH	Key Location/Qualifiers
FT	variation replace(581.C)

[illegible]

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	52.6	42.1	865	4	US-09-270-767-11042	Sequence 11042, A
C 2	51.4	41.1	670889	4	US-09-949-016-12505	Sequence 12505, A
3	51.4	41.1	670890	4	US-09-949-016-14207	Sequence 14207, A
4	50.6	40.5	152231	3	US-09-128-155-16	Sequence 16, Appl
5	50.2	40.2	1845	4	US-09-614-034-188	Sequence 188, App
C 6	49.8	39.8	204	4	US-09-107-433-2184	Sequence 2184, Ap
7	49.8	39.8	209	4	US-09-107-433-185	Sequence 185, App
8	49.8	39.8	282	4	US-09-107-433-184	Sequence 184, App
C 9	49.8	39.8	308	4	US-09-107-433-1723	Sequence 1723, Ap
10	49.8	39.8	612	4	US-09-107-433-1370	Sequence 1970, Ap
C 11	49.6	39.7	320	3	US-09-165-264-11	Sequence 11, Appl
12	49.4	39.5	318	3	US-09-165-264-12	Sequence 12, Appl
C 13	49.4	39.5	319	3	US-09-165-264-8	Sequence 8, Appli
14	49.4	39.5	320	3	US-09-165-264-7	Sequence 7, Appli
C 15	49.4	39.5	320	3	US-09-165-264-13	Sequence 13, Appl
16	49.4	39.5	320	3	US-09-165-264-14	Sequence 14, Appl
17	49	39.2	11766	4	US-09-949-016-12531	Sequence 12531, A
18	49	39.2	11770	4	US-09-949-016-12720	Sequence 12720, A
C 19	49	39.2	11770	4	US-09-949-016-13487	Sequence 13487, A
20	49	39.2	11770	4	US-09-949-016-13488	Sequence 13488, A
C 21	48.8	39.0	22206	4	US-09-543-681A-13901	Sequence 13901, A
22	47	37.6	336	4	US-09-543-681A-4150	Sequence 4150, Ap
C 23	47	37.6	339	4	US-09-543-681A-4147	Sequence 4147, Ap
24	47	37.6	483	4	US-09-543-681A-4138	Sequence 4138, Ap
C 25	47	37.6	516	4	US-09-543-681A-4140	Sequence 4140, Ap
26	47	37.6	519	4	US-09-543-681A-4145	Sequence 4145, Ap
C 27	47	37.6	608	4	US-09-543-681A-4115	Sequence 4115, Ap

RESULT 8

RESULT 8


```

; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

Query Match          39.7%; Score 49.6; DB 3; Length 320;
Best Local Similarity 67.3%; Pred.No.0.4;
Matches 70; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY      4 CTGCGCCCTGGCCCTGCCACCCACCCACCCACCCACCCACCCAGCGGCGCGCCGC 63
DB      304 CTGCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 245

QY      64 CGCGCCCGCGCGCGCGCGCGCTCGCGCGCACCACCGGTTCGCGCGC 107
DB      244 CCGCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 201

RESULT 12
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

Query Match          39.5%; Score 49.4; DB 3; Length 318;
Best Local Similarity 66.4%; Pred.No.0.43;
Matches 71; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY      1 CCCTCGCCCTGGCCCTGCCCCCACCCACCCACCCACCCACCCAGCGGCGCGCGC 60
DB      300 CCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 241

QY      61 GCCCGCCCCCGCGCGCGCGCGCTCGGCGCGACCCGGTTCCGCGC 107
DB      240 CCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 194

RESULT 13
US-09-165-264-8/c
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer sequence
US-09-165-264-8

Query Match          39.5%; Score 49.4; DB 3; Length 319;
Best Local Similarity 66.4%; Pred. No. 0.43;
Matches 71; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy      1  CCCCCTGCCCTGTGCCTCTGCCCCACACCCACCCACCCACCCACCCAGCGCGCGCCGC 60
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      300  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 241

Qy      61  GCCCGCCCCCGCGCGCGCGCGTGTGGCGCCGACCCCGGTTCCGCGCG 107
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      240  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 194
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 14
US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

```

```

RESULT 15
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinavagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

```

	Best Local Similarity	66.4%;	Pred. No.	0.43;	Mismatches	71;	Conservative	0;	Gaps	0;
QY	1	CCCCCTGCCCTGCCCACCCACCCACCCACCCACCCACCCACCCAGCGCGCGCCGC	60							
Db	300	CCC	241							
QY	61	GCCCGCCCCCGCGCGGGCGGGTCTGGCGCGAGCCCGGTTCGCGCGC	107							
Db	240	CCC	194							

Search completed: November 8, 2005, 19:52:44
Job time : 33.8505 secs

[illegible][illegible]

Qy 1 CCCCTGCCCTGCCCCCTGCCCCCAACCCCACCCCAGCGGGCCGC 60

Query match 43.8%
Best Local Similarity 67.5%
Matches 77; Conservative

Result No.	Score	Query %			DB	ID	Description
		Match	Length	Count			
C	1	65	52.0	821	8	AQ933474	HS_4832_A
	2	64.4	51.5	1283	9	AG335406	Mus_muscul
C	3	63.4	50.7	982	6	CD329933	AGENCOURT
	4	62.2	50.6	425	2	BE551555	hx97f12.x
C	5	62.4	49.9	1866	3	CR622226	full-length
	6	62.2	49.8	996	5	BO708087	AGENCOURT
C	7	62	49.6	1024	6	CA975437	AGENCOURT
	8	61.4	49.1	1674	9	CL078342	CH216-149
C	9	61.2	49.0	843	9	CNS00C51	AL059666
	10	61	48.8	506	9	CNS027Y	Drosophill
C	11	61	48.8	1003	9	CL475036	AL185259
	12	61	48.8	1285	9	AG334095	Tetraodon
C	13	60.8	48.6	908	9	CNS006B4	CL475036
	14	60.8	48.6	1134	5	BQ651163	Mus_muscul
C	15	60.4	48.3	840	9	AG043467	AL064031
	16	60.4	48.3	1101	9	CNS00153F	Drosophill
C	17	60.2	48.2	421	8	BH254057	AL104949
	18	60.2	48.2	566	9	CNS03JN0	BH254057
C	19	60.2	48.2	908	9	CL500140	SALIK_0159
	20	60.2	48.2	953	9	AG391895	AL247077
C	21	60.2	48.2	1014	9	CL492395	CL500140
	22	60.2	48.2	1046	5	BQ952254	SAIL_679
C	23	60.2	48.2	1065	5	BUI48615	AG391895
	24	60.2	48.2	1142	9	CL468510	Mus_muscul


```

DEFINITION hx97fi2.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3195791 3'
similar to contains element MSRI repetitive element ;, mRNA
sequence.
ACCESSION BE551555
VERSION BE551555.1 GI:9793338
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 425)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 206.
Location/Qualifiers
FEATURES
source
1..425
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3195791"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pYT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 50.6%; Score 63.2; DB 2; Length 425;
Best Local Similarity 68.8%; Pred. NO. 0.13;
Matches 86; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 CCCCTGCCCTGCCCCCTGCCCCACCCACCCACCCACCCACCCACCCACCCACCCACCCGCGCGCCGC 60
Dy 269 CCCCCCCCCCCCCCTCCCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCC 328
QY 61 GCGCGCCCCGCGCGCGCGCGCGCTCGCGCCGACCCCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 120
Dy 329 CCCCCCCCCCTCCCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 388
QY 121 CCAGC 125
Dy 389 CCGCG 393
RESULT 5
CR622226 LOCUS
DEFINITION CR622226 1866 bp mRNA linear HTC 21-JUL-2004
25-normalized cDNA clone CSODC004Y008 of Neuroblastoma Cot
ACCESSION CR622226
VERSION CR622226.1 GI:50503033
KEYWORDS HTC; CNSLT_cDNA.

```

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1866)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
REFERENCE 2 (bases 1 to 1866)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1866
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC004Y008"
/tissue_type="Neuroblastoma Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 49.9%; Score 62.4; DB 3; Length 1866;
Best Local Similarity 98.4%; Pred. NO. 0.13;
Matches 63; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 62 CCCCCCCCCGCGCGCGCGCGCTCGCGCCACCCGCTTCGCGCGCGCGCGCGCGCGCGCGCC 121
Dy 1 CCCCCCCCCGCGCGCGCGCGCTCGCGCCACCCGCTTCGCGCGCGCGCGCGCGCGCGCGCC 60
QY 122 CAGC 125
Dy 61 CAGC 64
RESULT 6
BO708087 LOCUS
DEFINITION BO708087 996 bp mRNA linear EST 16-JUL-2002
AGENCOURT_8475084 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301404
5', mRNA sequence.
ACCESSION BO708087
VERSION BO708087.1 GI:21846986
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health,
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LRCM2517 row: d column: 13
High quality sequence stop: 276.
Location/Qualifiers

```

source	1. .996 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6301404" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_113" /notes="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
Query Match	49.8%; Score 62.2; DB 5; Length 996;
Best Local Similarity	67.2%; Pred. No. 0.16;
Matches	82; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy	1 CCCTGCGCCCTGCGCCCGCCGCGCTCGGCGCCGACCGGTTCCGCGCGCGGAGCGGGGC 60 626 CCC 685
Db	
Qy	61 GCCCGCGCGCGCGCGCGCGCTCGGCGCCGACCGGTTCCGCGCGCGGAGCGGGGC 120 686 CCCCCGCC 745
Db	
Qy	121 CC 122
Db	746 GC 747
RESULT 7	
CA975437	
LOCUS	1024 bp mRNA linear EST 06-JAN-2003
DEFINITION	AGENCOURT 11067725 NCI CGAP ZEMB3 Danio rerio cDNA clone IMAGE:6803474 5', mRNA sequence.
ACCESSION	CA975437
VERSION	CA975437.1 GI:27508091
KEYWORDS	EST.
SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 1024)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. Chi-Bin Chien cDNA Library Preparation: Invitrogen Corp DNA sequencing by: The I.M.A.G.E. Consortium (LLNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLNL14326 row: h column: 01 High quality sequence start: 40 High quality sequence stop: 419. Location/Qualifiers 1. .1024 /organism="Danio rerio" /mol_type="mRNA" /db_xref="taxon:7955" /clone="IMAGE:6803474" /lab_host="DH10B (T1-resistant)" /tissue_type="embryo" /clone_lib="NCI CGAP ZEMB3" /notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
FEATURES	
source	
Query Match	49.1%; Score 61.4; DB 9; Length 1674;
Best Local Similarity	72.1%; Pred. No. 0.19;
Matches	80; Conservative 0; Mismatches 31; Indels 0; Gaps 0
Qy	1 CCCTGCGCCCTGCGCCCGCCGCGCTCGGCGCCGACCGGTTCCGCGCGGAGCGGGGC 60 190 CCCCCCGCCC 249
Db	
Qy	61 GCCCGCGCGCGCGCGCGCGCTCGGCGCCGACCGGTTCCGCGCGGAGCGGGGC 111
ORIGIN	
Query Match	49.6%; Score 62; DB 6; Length 1024;
Best Local Similarity	68.6%; Pred. No. 0.17;
Matches	83; Conservative 0; Mismatches 38; Indels 0; Gaps 0
Qy	1 CCCTGCGCCCTGCGCCCGCCGCGCTCGGCGCCGACCGGTTCCGCGCGGAGCGGGGC 60 598 CCCCCCGCCC 657
Db	
Qy	61 GCCCGCGCGCGCGCGCGCGCTCGGCGCCGACCGGTTCCGCGCGGAGCGGGGC 120 658 CCCCCCGCCNCCCGGCGCGCG 717
Db	
Qy	121 C 121
Db	718 C 718
RESULT 8	
CL078342	
LOCUS	1674 bp DNA linear GSS 31-DEC-2003
DEFINITION	CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-149D22, Genomic survey sequence.
ACCESSION	CL078342
VERSION	CL078342.1 GI:40534255
KEYWORDS	GSS.
SOURCE	Xenopus tropicalis (western clawed frog)
ORGANISM	Xenopus tropicalis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE	1 (bases 1 to 1674)
AUTHORS	Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE	A physical map of the xenopus tropicalis genome
JOURNAL	Unpublished (2003)
COMMENT	Contact: Richard K Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@wustl.edu Insert Length: 175000 Std Error: 0.00 Seq primer: Sp6 ATTAGTGACACTATAG Class: BAC ends High quality sequence start: 300 High quality sequence stop: 387. Location/Qualifiers 1. .1674 /organism="Xenopus tropicalis" /mol_type="genomic DNA" /strain="Nigerian frog" /db_xref="taxon:8364" /clone="CH216-149D22" /sex="male" /cell_line="Stock 248 F7A2, inbred N7" /clone_lib="CH216" /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
FEATURES	
source	
Query Match	49.1%; Score 61.4; DB 9; Length 1674;
Best Local Similarity	72.1%; Pred. No. 0.19;
Matches	80; Conservative 0; Mismatches 31; Indels 0; Gaps 0
Qy	1 CCCTGCGCCCTGCGCCCGCCGCGCTCGGCGCCGACCGGTTCCGCGCGGAGCGGGGC 60 190 CCCCCCGCCC 249
Db	
Qy	61 GCCCGCGCGCGCGCGCGCGCTCGGCGCCGACCGGTTCCGCGCGGAGCGGGGC 111

VERSION	AL185299.1	GI:7823403
KEYWORDS	GSS; genome survey sequence.	
SOURCE	Tetraodon nigroviridis	
ORGANISM	Tetraodon nigroviridis	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraontoidea; Tetraodontidae; Tetraodon.	
AUTHORS	Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Pizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.	
TITLE	Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence	
JOURNAL	Nat. Genet. 25 (2), 235-238 (2000)	
MEDLINE	20296633	
PUBMED	10835645	
REFERENCE		
AUTHORS	Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Pizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.	
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
JOURNAL	Genome Res. 10 (7), 939-949 (2000)	
MEDLINE	20359837	
PUBMED	10899143	
REFERENCE	3 (bases 1 to 506)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage	
COMMENT	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr) - Web : www.genoscope.cns.fr	
FEATURES	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.	
source	Location/Qualifiers	
1..506	/organism="Tetraodon nigroviridis"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:99883"	
	/clone="243J12"	
	/clone_lib="G"	
	/notes="Genoscope sequence ID : COAG243D806SP1-end : PUC-Ori"	
ORIGIN		
Query Match	48.8%; Score 61; DB 9; Length 506;	
Best Local Similarity	52.1%; Pred. No. 0.29;	
Matches	62; Conservative 30; Mismatches 27; Indels 0; Gaps 0;	
Qy	1 CCCCTGCGCCCTGCGCCCGCGCGCTGCGCCGACCGGTTCCGCGCGGCGAGCGCGG 60	
Db	344 CC 285	
Qy	61 GCCCGCGCCCGCGCGCGCGCTGCGCCGACCGGTTCCGCGCGGCGAGCGCGG 119	
Db	284 CCCCSC 226	
RESULT 11		
LOCUS	CL475036	1003 bp DNA linear GSS 01-APR-2004
DEFINITION	SAIL_230_F07.v1 SAIL Collection Arabidopsis thaliana genomic clone	
ACCESSION	SAIL_230_F07.v1, genomic survey sequence.	
VERSION	CL475036	
KEYWORDS	CL475036.1 GI:45940747	
SOURCE	GSS.	
ORGANISM	Arabidopsis thaliana (thale cress)	
	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:52:53 ; Search time 4078.01 Seconds
(without alignments)
9232.381 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144

Perfect score: 777
Sequence: 1 tcattgttagtaagactgtgt.....gtctctttttttattgttaag 777

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pl.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	765.8	98.6	135675	9 AC093174	Homo sapi
2	765.8	98.6	148828	9 AX157024	Homo sapi
3	765.8	98.6	180963	9 AC091492	Homo sapi
C 4	765.8	98.6	221402	2 AC015995	Homo sapi
5	764.2	98.4	851	9 AF548352	Homo sapi
C 6	565.4	72.8	186028	2 AC151849	Callithri
7	421	54.2	65160	2 AC135179	Homo sapi
8	130.6	16.8	212216	2 AC131898	Oryctolag
9	103.6	13.3	185737	2 AC125447	Mus muscu
C 10	88.6	11.4	191540	2 AC136055	Rattus no
C 11	88.6	11.4	245724	2 AC091418	Rattus no
C 12	88.6	11.4	250169	2 AC120668	Rattus no
13	68.6	8.8	154195	2 AC142242	Ateles
14	53	6.8	183103	5 BX465190	Zebrafish
15	52.8	6.8	145160	2 CR847935	Danio rer
16	51.8	6.7	166021	2 AC117751	Mus muscu
C 17	50.8	6.5	153477	2 AC06278	Plasmodiu
C 18	50.8	6.5	251551	3 AB014844	Plasmodiu
19	50	6.4	5771	6 AX344664	Sequence

20	50	6.4	125020	9 AF429315	Homo sapi
C 21	49.6	6.4	95996	8 ATFF18P9	Arabidops
22	49.4	6.4	175188	2 BX897715	Danio rer
23	49.2	6.3	18683	6 AX281291	Sequence
24	49.2	6.3	18683	6 AX345214	Sequence
C 25	49.2	6.3	156276	2 CR769765	Danio rer
C 26	49.2	6.3	175234	5 BX547997	Zebrafish
C 27	49.2	6.3	241364	2 BX571898	Danio rer
28	49	6.3	114222	6 AX345121	Sequence
29	49	6.3	114222	6 AX348323	Sequence
30	49	6.3	138121	9 AC114806	Homo sapi
31	48.8	6.3	110000	2 PFMA18P1_08	Continuation (9 of
32	48.8	6.3	163443	2 AC006280	Plasmodiu
33	48.8	6.3	196149	2 AC004709	Plasmodiu
C 34	48.8	6.3	252650	3 AE014847	Plasmodiu
35	48.6	6.3	179824	9 AC096708	Homo sapi
36	48.4	6.2	13123	6 AX281322	Sequence
37	48.4	6.2	13123	6 AX344419	Sequence
38	48	6.2	92725	9 AC010350	Homo sapi
39	48	6.2	152706	9 AC008568	Homo sapi
C 40	48	6.2	152707	9 AC008876	Homo sapi
41	48	6.2	349980	6 AX344565	Sequence
42	47.6	6.1	958	3 NAM556112	Necator a
43	47.6	6.1	958	3 NAM556113	Necator a
44	47.6	6.1	958	3 NAM556114	Necator a
45	47.6	6.1	958	3 NAM556116	Necator a

ALIGNMENTS

RESULT 1
AC093174
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-167M22 map 3p, complete
sequence.
AC093174
VERSION AC093174.2 GI:24796717
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 135675)
AUTHORS
Wu, Q., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,
Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,
Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,
Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,
Liu, Y., Li, W., Li, W., Li, X., Luo, J., Luo, Y., Qi, Q., Qi, X., Song, L.,
Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,
Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,
Wang, Y., Wu, D., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,
Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,
Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
Yu, J. and Yang, H.

Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 135675)

REFERENCE 2 (bases 1 to 135675)
AUTHORS
Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,
He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,
Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,
Li, W., Li, W., Li, X., Luo, J., Luo, Y., Qi, Q., Qi, X., Song, L.,
Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,
Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y.,
Wu, D., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,
Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,
Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
Yu, J. and Yang, H.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (13-AUG-2001) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China

```

3 (bases 1 to 135675)
Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
Guo,Z., He,L., Hu,S., Huang,F., Jin,F., Kang,N., Li,C., Li,C.,
Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Liu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y.,
Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X.,
Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and
Yang,H.
Direct Submission
Submitted (08-NOV-2002) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Nov 8, 2002 this sequence version replaced gi:15148930.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:Hgc@igtp.ac.cn
-----Project Information
Center project name:1k project
Center clone name: RP11-167M22
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; Et 55% of reads
Assembly program: Big Dye; 45% of reads
Consensus quality: Phrap; version 0.990329
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 6 bases at least Q20
Insert size: 3392; sum-of-contigs
Quality coverage: 0.00x in Q20 bases;sum-of-contigs
-----
FEATURES             source
Source               1. 135675
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="3"
                        /map="3p"
                        /clone="RP11-167M22"
ORIGIN
Query Match          98.6%; Score 765.8; DB 9; Length 135675;
Best Local Similarity 99.1%; Pred. No. 6.5e-160;
Matches 770; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 TCATGTAGTAAAGCTGTAGATATCGGGTCTCGATGTGGCGCTATTCAAGCCCTGA 60
Db 41828 TCATGTAGTAAAGCTGTAGATATCGGGTCTCGATGTGGCGCTATTCAAGCCCTGA 41987
Qy 61 TGATAAGGCTTTTGGCATTAGATCGTGTGTTTGTCTTCATGGAATAACAGCTATCTAGG 120
Db 41988 TGATAAGGCTTTTGGCATTAGATCGTGTGTTTGTCTTCATGGAATAACAGCTATCTAGG 41947
Qy 121 ATCTTGAAGCCTTTTCATAAGAGATAAGGTTGTGAATCCTTAAGACCCCTAGGACCAATTACT 180
Db 41948 ATCTTGAAGCCTTTTCATAAGAGATAAGGTTGTGAATCCTTAAGACCCCTAGGACCAATTACT 42007
Qy 181 TAGATGATCTGCTCTCTGGTTCGTCCTGAAAGTCTGCTTCTGTAGGGGTGCTGCA 240
Db 42008 TAGATGATCTGCTCTCTGGTTCGTCCTGAAAGTCTGCTTCTGTAGGGGTGCTGCA 42067
Qy 241 TTTGCCCTTGCCTAAGTGGTGGGACACAACTGACTGTACCTTAGGCTTTAATAACCAT 300
Db 42068 TTTGCCCTTGCCTAAGTGGTGGGACACAACTGACTGTACCTTAGGCTTTAATAACCAT 42127
Qy 301 GTGTCATCTAGATGAAGTTATATTTTAAAGAGATCGTTTTTGGCATGTATATAATTTTC 360

```

```

42128 GTGTCATCTAGATGAAGTTATATTTTACTAAGGATCGTTTTTGGCATGTATAACTTCTC 42187
Qy 361 AAACATTAACTTTTCAGGGTTATTAAATCCTTTTAAGGCTAGTTTTTCTTAAGTCTGTGCA 420
Db 42188 AAACATTAACTTTTCAGGGTTATTAAATCCTTTTAAGGCTAGTTTTTCTTAAGTCTGTGCA 42247
Qy 421 GTAATAGAGGTATCGTCATTTCATCATGTGACATATAAAGATGGAAGGGGCTTCATTATGTTA 480
Db 42248 GTAATAGAGGTATCGTCATTTCATCATGTGACATATAAAGATGGAAGGGGCTTCATTATGTTA 42307
Qy 481 GTGATGAAATAGGAAAGTAGGTAAGTATTTTAAATAGATGTTTCTTTTATGAATAAT 540
Db 42308 GTGATGAAATAGGAAAGTAGGTAAGTATTTTAAATAGATGTTTCTTTTATGAATAAT 42367
Qy 541 TTTTAAAGATTGTCAGCCCTGCATCATTTATGATGAATCATTTTGGTCTGTAGTT 600
Db 42368 TTTTAAAGATTGTCAGCCCTGCATCATTTATGATGAATCATTTTGGTCTGTAGTT 42427
Qy 601 ACTTTTAGAGATAGAAAGCATTTGTAGGCTCAGGGAAGCAACATTCAGAAATGAAATCC 660
Db 42428 ACTTTTAGAGATAGAAAGCATTTGTAGGCTCAGGGAAGCAACATTCAGAAATGAAATCC 42487
Qy 661 AATAGAGAAGGTAAATTTATTTGGCATGTACATTTTGGCAGCCTAGCTGTGACATGT 720
Db 42488 AATAGAGAAGGTAAATTTATTTGGCATGTACATTTTGGCAGCCTAGCTGTGACATGT 42547
Qy 721 GTACACATCTGACATGTGTATATTTGAAATCTTGTCTCTTTTATTTGTTAAG 777
Db 42548 GTACACATCTGACATGTGTATATTTGAAATCTTGTCTCTTTTATTTGTTAAG 42604

RESULT 2
AY157024 148828 bp DNA linear PRI 16-JUN-2004
LOCUS Homo sapiens peroxisome proliferative activated receptor gamma
DEFINITION (PPARG) gene, complete cds.
ACCESSION AY157024
VERSION AY157024.2 GI:48762804
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148828)
Rieder,M.J., da Ponte,S.H., Kuldane,K., Rajkumar,N., Smith,J.D.,
Toth,E.J. and Nickerson,D.A.
Direct Submission
Submitted (30-SEP-2002) Genome Sciences, University of Washington,
1705 NE Pacific, Seattle, WA 98195, USA
REFERENCE 2 (bases 1 to 148828)
Rieder,M.J., Daniels,R.L., da Ponte,S.H., Hastings,N.C.,
Ahearn,M.O., Rajkumar,N., Yi,Q. and Nickerson,D.A.
Direct Submission
Submitted (16-JUN-2004) Genome Sciences, University of Washington,
1705 NE Pacific, Seattle, WA 98195, USA
REMARK Sequence update by submitter
COMMENT On Jun 16, 2004 this sequence version replaced gi:23953882.
To cite this work please use: SeattleSNPs. NHLBI Program for
Genomic Applications, UW-FHCRC, Seattle, WA
(URL:http://pga.mbt.washington.edu).
FEATURES             Location/Qualifiers
source               1. 148828
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
repeat_region       9. 59
                        /rpt_family="L1"
variation           54
                        /frequency="0.12"
                        /replace="g"
repeat_region       60. 351
                        /rpt_family="Alu"

```



```
/rpt family="L2"
/rpt_type=dispersed
22205..22492
/rpt_family="alu"
/rpt_type=dispersed
22493..22581
/rpt_family="L2"
/rpt_type=dispersed
22582..22916
/rpt_family="MER1_type"
/rpt_type=dispersed
22917..23104
/rpt_family="L2"
/rpt_type=dispersed

Query Match 98.6%; Score 765.8; DB 9; Length 148828;
Best Local Similarity 99.1%; Pred. No. 6.5e-160;
Matches 770; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TCATGTAGTAAAGTGTAGTAAAGTGTGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
Db 25723 TCATGTAGTAAAGTGTAGTAAAGTGTGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 25782

Qy 61 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTCATGGAAAAATACAGCTATTCTAGG 120
Db 25783 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTCATGGAAAAATACAGCTATTCTAGG 25842

Qy 121 ATCTTTGAGCCCTTTTCATAGAGATAGTGTGTAATCTTAAGACCTAGGACCAATTACT 180
Db 25843 ATCTTTGAGCCCTTTTCATAGAGATAGTGTGTAATCTTAAGACCTAGGACCAATTACT 25902

Qy 181 TAGATGATCTCTCTCTGTTTCTGCTCTGAAAGTCTGCTTCTGTAGGGGTGTGCTGCA 240
Db 25903 TAGATGATCTCTCTCTGTTTCTGCTCTGAAAGTCTGCTTCTGTAGGGGTGTGCTGCA 25962

Qy 241 TTTGCCCTTGCCTAAGTGTGTGGCACACAACTGTACTGTCTACCTTAGGCTTTAATAACCAT 300
Db 25963 TTTGCCCTTGCCTAAGTGTGTGGCACACAACTGTACTGTCTACCTTAGGCTTTAATAACCAT 26022

Qy 301 GTGTCATCTAGAAAGTGTATATTTTAAAGAGATGCTTTTCCCATGTATATAATTTTC 360
Db 26023 GTGTCATCTAGAAAGTGTATATTTTAAAGAGATGCTTTTCCCATGTATATAATTTTC 26082

Qy 361 AAACATTAACCTTCAGGGTATTAATCTTTAAGGCTCTAGTTTCTTAAAGTCTGTGCA 420
Db 26083 AAACATTAACCTTCAGGGTATTAATCTTTAAGGCTCTAGTTTCTTAAAGTCTGTGCA 26142

Qy 421 GTAATAGAGGTATCTCATTCATGTGACATATAAGATGGAAGGGGCTTCATTCATGTTA 480
Db 26143 GTAATAGAGGTATCTCATTCATGTGACATATAAGATGGAAGGGGCTTCATTCATGTTA 26202

Qy 481 GTGATGGAATAGGAAGTAGTGAAGTATTTTAATAGATGTTTCTTTATGAAATAAT 540
Db 26203 GTGATGGAATAGGAAGTAGTGAAGTATTTTAATAGATGTTTCTTTATGAAATAAT 26262

Qy 541 TTTTAAAGATGTGTCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTAGTT 600
Db 26263 TTTTAAAGATGTGTCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTAGTT 26322

Qy 601 ACTTTTATAGAAATAGAAAGCATTTAGGCTCAGGGAAGCAAACTCAGAAATCAATCC 660
Db 26323 ACTTTTATAGAAATAGAAAGCATTTAGGCTCAGGGAAGCAAACTCAGAAATCAATCC 26382

Qy 661 AATAGAGAGTAAATTTATTTGGGCATGTACATTTTGGCAGCCCTAGGCTGTGACATGT 720
Db 26383 AATAGAGAGTAAATTTATTTGGGCATGTACATTTTGGCAGCCCTAGGCTGTGACATGT 26442

Qy 721 GTACACATTCGAACATGTGTATATTTGAAATCTTGTCTCTCTTTTATTTGTTAAG 777
Db 26443 GTACACATTCGAACATGTGTATATTTGAAATCTTGTCTCTCTTTTATTTGTTAAG 26499
```

AC091492	AC091492	180963 bp	DNA	linear	PRI 20-DEC-2002
LOCUS	Homo sapiens chromosome 3 clone RP11-33519 map 3p, complete				
DEFINITION	Homo sapiens chromosome 3 clone RP11-33519 map 3p, complete				
AC091492	AC091492	180963 bp	DNA	linear	PRI 20-DEC-2002
VERSION	AC091492.2	GI:24796732			
KEYWORDS	HTG				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 180963)				
	Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
TITLE	Chromosome 3p genomic sequence				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 180963)				
AUTHORS	Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-APR-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China				
REFERENCE	3 (bases 1 to 180963)				
AUTHORS	Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-NOV-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China				
REFERENCE	4 (bases 1 to 180963)				
AUTHORS	Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-NOV-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China				
REFERENCE	4 (bases 1 to 180963)				
AUTHORS	Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-DEC-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China				

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 5, 2000 this sequence version replaced gi:7329342.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 12392
 Center clone name: 335_1_9
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 97% of reads
 Sequencing vector: Plasmid; n/a; %0.f% of reads
 2.51630941286114Chemistry: Dye-terminator Big Dye; 100% of reads

 Assembly program: Phrap; version 0.960731
 Consensus quality: 188096 bases at least Q40
 Consensus quality: 198983 bases at least Q30
 Consensus quality: 206638 bases at least Q20
 Insert size: 182000; agarose-fp
 Insert size: 217902; sum-of-contigs
 Quality coverage: 6.0 in Q20 bases; agarose-fp
 Quality cov.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 36 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 3984: contig of 3984 bp in length
 * 3985 4084: gap of 100 bp
 * 4085 5217: contig of 1133 bp in length
 * 5218 5317: gap of 100 bp
 * 5318 6422: contig of 1105 bp in length
 * 6423 6522: gap of 100 bp
 * 6523 7602: contig of 1080 bp in length
 * 7603 7702: gap of 100 bp
 * 7703 8882: contig of 1180 bp in length
 * 8883 8982: gap of 100 bp
 * 8983 10209: contig of 1227 bp in length
 * 10210 10309: gap of 100 bp
 * 10310 12072: contig of 1763 bp in length
 * 12073 12172: gap of 100 bp
 * 12173 13428: contig of 1256 bp in length
 * 13429 13528: gap of 100 bp
 * 13529 14796: contig of 1268 bp in length
 * 14797 14896: gap of 100 bp
 * 14897 16049: contig of 1153 bp in length
 * 16050 16149: gap of 100 bp
 * 16150 17583: contig of 1434 bp in length
 * 17584 17683: gap of 100 bp
 * 17684 18808: contig of 1125 bp in length
 * 18809 18908: gap of 100 bp
 * 18909 20356: contig of 1448 bp in length
 * 20357 20456: gap of 100 bp
 * 20457 22555: contig of 2099 bp in length
 * 22556 22655: gap of 100 bp
 * 22656 24625: contig of 1970 bp in length
 * 24626 24725: gap of 100 bp
 * 24726 26352: contig of 1627 bp in length
 * 26353 26452: gap of 100 bp
 * 26453 28366: contig of 1914 bp in length
 * 28367 28466: gap of 100 bp
 * 28467 29704: contig of 1238 bp in length
 * 29705 29804: gap of 100 bp
 * 29805 31337: contig of 1533 bp in length
 * 31338 31437: gap of 100 bp
 * 31438 32941: contig of 1504 bp in length

32942 33041: gap of 100 bp
 * 33042 34690: contig of 1649 bp in length
 * 34691 34790: gap of 100 bp
 * 34791 36510: contig of 1720 bp in length
 * 36511 36610: gap of 100 bp
 * 36611 39065: contig of 2455 bp in length
 * 39066 39165: gap of 100 bp
 * 39166 41551: contig of 2386 bp in length
 * 41552 41651: gap of 100 bp
 * 41652 44456: contig of 2805 bp in length
 * 44457 44556: gap of 100 bp
 * 44557 46379: contig of 1823 bp in length
 * 46380 46479: gap of 100 bp
 * 46480 49222: contig of 2743 bp in length
 * 49223 49322: gap of 100 bp
 * 49323 52311: contig of 2989 bp in length
 * 52312 52411: gap of 100 bp
 * 52412 57964: contig of 5553 bp in length
 * 57965 58064: gap of 100 bp
 * 58065 63519: contig of 5455 bp in length
 * 63520 63619: gap of 100 bp
 * 63620 70598: contig of 6979 bp in length
 * 70599 70698: gap of 100 bp
 * 70699 83449: contig of 12751 bp in length
 * 83450 83549: gap of 100 bp
 * 83550 100089: contig of 16540 bp in length
 * 100090 100189: gap of 100 bp
 * 100190 120532: contig of 20343 bp in length
 * 120533 120632: gap of 100 bp
 * 120633 160594: contig of 39962 bp in length
 * 160595 160694: gap of 100 bp
 * 160695 221402: contig of 60708 bp in length.
 *
 FEATURES
 Location/Qualifiers
 1..221402
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3"
 /clone="RP11-335I9"
 /clone_lib="RPC1-11 Human Male BAC"
 1..3984
 /note="assembly_fragment
 clone end:SP6
 vector side:left"
 4085..5217
 /note="assembly_fragment"
 5318..6422
 /note="assembly_fragment"
 6523..7602
 /note="assembly_fragment"
 7703..8882
 /note="assembly_fragment"
 8983..10209
 /note="assembly_fragment"
 10310..12072
 /note="assembly_fragment"
 12173..13428
 /note="assembly_fragment"
 13529..14796
 /note="assembly_fragment"
 14897..16049
 /note="assembly_fragment"
 16150..17583
 /note="assembly_fragment"
 17684..18808
 /note="assembly_fragment"
 18909..20356
 /note="assembly_fragment"
 20457..22555
 /note="assembly_fragment"
 Query Match 98.6%; Score 765.8; DB 2; Length 221402;

```
Best Local Similarity 99.1%; Pred. No. 6.1e-160;
Matches 770; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TCATGTAGCTAAGACTGTGTAGAACTGCGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
Db 98593 TCATGTAGCTAAGACTGTGTAGAACTGCGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 98534
Qy 61 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTTCATGAAGAAATACAGCTATTCTAGG 120
Db 98533 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTTCATGAAGAAATACAGCTATTCTAGG 98474
Qy 121 ATCTTGGAGCTTTTCATAGAGATAAGCTTCTGAATCCTAAGACCTAGGACCATTTTACT 180
Db 98473 ATCTTGGAGCTTTTCATAGAGATAAGCTTCTGAATCCTAAGACCTAGGACCATTTTACT 98414
Qy 181 TAGATGATCTGCTCTCTGTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
Db 98413 TAGATGATCTGCTCTCTGTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 98354
Qy 241 TTTCCTTGGCCTAAGTGTGTGGGCAACAACCTGCTACCTTGGCTTTAATAACCAT 300
Db 98353 TTTCCTTGGCCTAAGTGTGTGGGCAACAACCTGCTACCTTGGCTTTAATAACCAT 98294
Qy 301 GTGTCATCTAGAACTGATATATTTTAAAGGATCGTTTGGCCATGTATATTTTC 360
Db 98293 GTGTCATCTAGAACTGATATATTTTAAAGGATCGTTTGGCCATGTATATTTTC 98234
Qy 361 AAACATTAACTTTTCAAGGCTTATTAATCCTTTTAAAGGCTAGTTTCTTAAAGTCTGTGCA 420
Db 98233 AAACATTAACTTTTCAAGGCTTATTAATCCTTTTAAAGGCTAGTTTCTTAAAGTCTGTGCA 98174
Qy 421 GTAATAGAGGTATCGTCAATTCATGTGACATATAAGATGGAAGGGCTTCATTTCATGTTA 480
Db 98173 GTAATAGAGGTATCGTCAATTCATGTGACATATAAGATGGAAGGGCTTCATTTCATGTTA 98114
Qy 481 GTGATGGAATAGGAAGCTAGTGAAGTATTTTAAATAGATGTTTCTTTTATGAATAAT 540
Db 98113 GTGATGGAATAGGAAGCTAGTGAAGTATTTTAAATAGATGTTTCTTTTATGAATAAT 98054
Qy 541 TTTTAAAGATTTGTCAGCCCTGCAATGATTTATGATGAATCAATTTTGGTCTGTAGTT 600
Db 98053 TTTTAAAGATTTTCCAGCCCTGCAATGATTTATGATGAATCAATTTTGGTCTGTAGTT 97994
Qy 601 ACTTTAGAGATAGAAAGCAATTTAGGCTCAGGAAAGCAAACTCAGAAATGAATCC 660
Db 97993 ACTTTAGAGATAGAAAGCAATTTAGGCTCAGGAAAGCAAACTCAGAAATGAATCC 97934
Qy 661 AATAGAGAGCTTAATTTTATTTGGCATGTACATTTTGGCAGCCCTAGGCTGTACATGT 720
Db 97933 AATAGAGAGCTTAATTTTATTTGGCATGTACATTTTGGCAGCCCTAGGCTGTACATGT 97874
Qy 721 GTACACATTTGAAATGATGTATATTTGAAATCTGCTCTCTTTTATTTGTTAAG 777
Db 97873 GTACACATTTGAAATGATGTATATTTGAAATCTGCTCTCTTTTATTTGTTAAG 97817

RESULT 5
AP548352
LOCUS
DEFINITION Homo sapiens peroxisome proliferator-activated receptor gamma 3
AP548352
ACCESSION
VERSION AP548352.1 GI:28626258
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 851)
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Auwerx, J. and Amouyel, P.
Meirhaeghe, A., Fajaa, L., Helbecque, N., Cottel, D., Helbecque, N.,
Auwerx, J. and Amouyel, P.
A functional polymorphism in a STAT5B site of the human PPAR gamma
3 gene promoter affects height and lipid metabolism in a French
```

```
population
JOURNAL Arterioscler. Thromb. Vasc. Biol. (2003) In press
REFERENCE 2 (bases 1 to 851)
AUTHORS Meirhaeghe, A., Fajaa, L., Helbecque, N., Auwerx, J. and Amouyel, P.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-2002) INSERM U508, Institut Pasteur de Lille, 1
rue du Pr Calmette - BP 245, Lille 59019, France
FEATURES
source
1..851
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p25"
<1..>851
/gene="PPARG"
/feature="PPARG"
/promoter
<1..777
/feature="PPARG"
/feature="promoter 3"
97
/feature="PPARG"
/feature="SNP"
/replace="g"
225
/feature="PPARG"
/feature="SNP"
/replace="c"
778..>851
/gene="PPARG"
/product="peroxisome proliferator-activated receptor gamma
3"
778..851
/feature="PPARG"
/number=42

ORIGIN
Query Match 98.4%; Score 764.2; DB 9; Length 851;
Best Local Similarity 99.0%; Pred. No. 3.1e-159;
Matches 769; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 TCATGTAGCTAAGACTGTGTAGAACTGCGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
Db 1 TCATGTAGCTAAGACTGTGTAGAACTGCGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
Qy 61 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTTCATGGAATAACAGCTATTCTAGG 120
Db 61 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTTCATGGAATAACAGCTATTCTAGG 120
Qy 121 ATCTTGGAGCTTTTCATAGAGATAAGCTTCTGAATCCTAAGACCTAGGACCATTTTACT 180
Db 121 ATCTTGGAGCTTTTCATAGAGATAAGCTTCTGAATCCTAAGACCTAGGACCATTTTACT 180
Qy 181 TAGATGATCTGCTCTCTGTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
Db 181 TAGATGATCTGCTCTCTGTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
Qy 241 TTTCCTTGGCCTAAGTGTGTGGGCAACAACCTGCTACCTTGGCTTTAATAACCAT 300
Db 241 TTTCCTTGGCCTAAGTGTGTGGGCAACAACCTGCTACCTTGGCTTTAATAACCAT 300
Qy 301 GTGTCATCTAGAACTGATATATTTTAAAGGATCGTTTGGCCATGTATATTTTC 360
Db 301 GTGTCATCTAGAACTGATATATTTTAAAGGATCGTTTGGCCATGTATATTTTC 360
Qy 361 AAACATTAACTTTTCAAGGCTTATTAATCCTTTTAAAGGCTAGTTTCTTAAAGTCTGTGCA 420
Db 361 AAACATTAACTTTTCAAGGCTTATTAATCCTTTTAAAGGCTAGTTTCTTAAAGTCTGTGCA 420
Qy 421 GTAATAGAGGTATCGTCAATTCATGTGACATATAAGATGGAAGGGCTTCATTTCATGTTA 480
Db 421 GTAATAGAGGTATCGTCAATTCATGTGACATATAAGATGGAAGGGCTTCATTTCATGTTA 480
```


AC135179
LOCUS Homo sapiens chromosome 17 clone RP11-1200B1 map 17, LOW-PASS linear HTG 10-OCT-2002
SEQUENCE SAMPLING.
AC135179
VERSION AC135179.2 GI:23683246
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 65160)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Birren, B., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 17, clone RP11-1200B1
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 65160)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,
Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tsefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (08-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 65160)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,
Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tsefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (10-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On Oct 10, 2002 this sequence version replaced gi:23592138.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28344
Center clone name: 1200_B_1

* NOTE: This record contains 80 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 720: contig of 720 bp in length
* 721 820: gap of 100 bp
* 821 1524: contig of 704 bp in length
* 1525 2325: contig of 701 bp in length
* 1625 2425: gap of 100 bp
* 2326 3211: contig of 786 bp in length
* 3212 3311: gap of 100 bp
* 3312 4038: contig of 727 bp in length
* 4039 4138: gap of 100 bp
* 4139 4869: contig of 731 bp in length
* 4870 4969: gap of 100 bp
* 4970 5704: contig of 735 bp in length
* 5705 5804: gap of 100 bp
* 5805 6538: contig of 734 bp in length
* 6539 6638: gap of 100 bp
* 6639 7365: contig of 727 bp in length
* 7366 7465: gap of 100 bp
* 7466 8184: contig of 719 bp in length
* 8185 8284: gap of 100 bp
* 8285 8988: contig of 704 bp in length
* 8989 9088: gap of 100 bp
* 9089 9806: contig of 718 bp in length
* 9807 10615: contig of 709 bp in length
* 10616 10715: gap of 100 bp
* 10716 11447: contig of 732 bp in length
* 11448 11547: gap of 100 bp
* 11548 12264: contig of 717 bp in length
* 12265 12364: gap of 100 bp
* 12365 13073: contig of 709 bp in length
* 13074 13173: gap of 100 bp
* 13174 13891: contig of 718 bp in length
* 13892 13991: gap of 100 bp
* 13992 14714: contig of 723 bp in length
* 14715 14814: gap of 100 bp
* 14815 15516: contig of 702 bp in length
* 15517 15616: gap of 100 bp
* 15617 16335: contig of 719 bp in length
* 16336 16435: gap of 100 bp
* 16436 17152: contig of 717 bp in length
* 17153 17252: gap of 100 bp
* 17253 17967: contig of 715 bp in length
* 17968 18067: gap of 100 bp
* 18068 18788: contig of 721 bp in length
* 18789 18888: gap of 100 bp
* 18889 19629: contig of 741 bp in length
* 19630 19730: gap of 100 bp
* 19730 20457: contig of 728 bp in length
* 20458 20557: gap of 100 bp
* 20558 21286: contig of 729 bp in length
* 21287 21386: gap of 100 bp
* 21387 22110: contig of 724 bp in length
* 22111 22922: contig of 712 bp in length
* 22923 23022: gap of 100 bp
* 23023 23444: contig of 722 bp in length
* 23445 23845: gap of 100 bp
* 23845 24675: gap of 100 bp
* 24676 25397: contig of 722 bp in length
* 25398 25497: gap of 100 bp
* 25498 26104: contig of 607 bp in length
* 26105 26104: contig of 607 bp in length


```
* 9217 9316: gap of unknown length
* 9317 15539: contig of 6223 bp in length
* 15540 15539: gap of unknown length
* 15640 21088: contig of 5449 bp in length
* 21089 21088: gap of unknown length
* 21189 33983: contig of 12795 bp in length
* 33984 34083: gap of unknown length
* 34084 48093: contig of 14010 bp in length
* 48094 48193: gap of unknown length
* 48194 67641: contig of 19448 bp in length
* 67642 91391: contig of 23650 bp in length
* 91392 91491: gap of unknown length
* 91492 117177: contig of 25686 bp in length
* 117178 117277: gap of unknown length
* 117278 144157: contig of 26880 bp in length
* 144158 144257: gap of unknown length
* 144258 212216: contig of 67959 bp in length.
FEATURES             Location/Qualifiers
     source           1..212216
                     /organism="Oryctolagus cuniculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9986"
                     /clone="LBI-83W7"
ORIGIN
Query Match          16.8%; Score 130.6; DB 2; Length 212216;
Best Local Similarity 62.5%; Pred. No. 4.8e-19;
Matches 307; Conservative 0; Mismatches 144; Indels 40; Gaps 5;
QY 287 GCCTTAATACCATGTCATCTAGAAAGTATATATTTTAAAGAGTCGTTTTGCC 346
Db 84838 GCGTTAAACACCCACCTCGCGCTGGAATCAATTTATA-TTCAGGAACATATATTTCTATGAC 84896
QY 347 ATGTATATAATTTTCAACATTAACCTTCAGGGTATTAACTCTTTTAAAGGCTAGTTTT 406
Db 84897 TTTCTGTCATC-----TGCCTTTCCAGTATTAATCTCTTTTAAAGATCTAGTTTT 84950
QY 407 CTTAAGTCGTGCAGTAATAGAGGTATCGTCAATTCATGTGACATATAAAGATGGAAGGGG 466
Db 84951 CTTAAGTCGCAATAATATATATGTATCAGTATTAATTTGGCATATAAAGATGGATGAT 85010
QY 467 CTTCAATTCATGTAGTAGTGAATAGGAATAGGAATAGGATGATTTTAAATAGATGTTTC 526
Db 85011 AAAGACCTTCGACTCACAATAGTCACTCACTGCACTGGAATAAAGCAATCAATTCATGT 85070
QY 527 TTTTATGAATAATTTTAAAGATTGTCAGCCCTGCGATGATTTATGATCAATCATTTT 586
Db 85071 CTCATAAAATCATTTTAAAGAGATTTTGCATCCCTGCTGTGATTTATGATGAGTCACCTGG 85130
QY 587 GTGGTCTGTTAGTTACTTTTATAGAAATAGAAAGCATTTGAGGCTCAGGGAAGCAAAACAT 646
Db 85131 GTGGCTCTGTTGTTACTTTCTAGAACAAAG-----CAT 85166
QY 647 TCAGAAATGAATCCAAATAGAGAGTAATTTATTTGGGCAATGATACATTTTGGCAGCCTA 706
Db 85167 TCAGAAATGAATCCAAACAAAGCA-AAAAATTTATTTGATGATGATTTTGGCAACCTA 85225
QY 707 GGCTGTGTACATGCTACACATCTTCGACATGCTGTATATTTGAAATCTGTCTCTTTT 766
Db 85226 G-----ATTGCTAAGGATTTTAAATATGTGTATATTTGAACATTTTATTTCTTTT 85277
QY 767 TTATGTTAAG 777
Db 85278 GAAATGTTTCAG 85288
RESULT 9
AC125447
LOCUS
DEFINITION Mus musculus chromosome 6 clone RP24-507D15 map 6, *** SEQUENCING
IN PROGRESS ***, 3 unordered pieces.
ACCESSION AC125447
```

VERSION
KEYWORDS
SOURCE
ORGANISMAC125447.4 GI:45237223
HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
Mus musculus (house mouse)

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Bixen, B., Nusbaum, C. and Lander, E.

TITLE

Mus musculus chromosome 6, clone RP24-507D15

REFERENCE

Unpublished

AUTHORS

2 (bases 1 to 185737)
Bixen, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 185737)

AUTHORS

Bixen, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hages, B., Galagan, J., Gardyna, S.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (06-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Mar 6, 2004 this sequence version replaced gi:29029334.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26714
Center clone name: 507_D_15-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 54649: contig of 54649 bp in length
 * 54650 54749: gap of 100 bp
 * 54750 156594: contig of 101845 bp in length
 * 156595 156694: gap of 100 bp
 * 156695 185737: contig of 23043 bp in length.

FEATURES

Location/Qualifiers
 1. 185737
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="6"
 /map="6"
 /clone="RP24-507D15"
 /clone_lib="RPC1-24 Male Mouse BAC"

ORIGIN

```

Query Match      13.3%; Score 103.6; DB 2; Length 185737;
Best Local Similarity 57.7%; Pred. No. 4.8e-13;
Matches 225; Conservative 0; Mismatches 159; Indels 6; Gaps 2;

Qy 371 TTTCCAGGGTTATTATCCCTTTTAAGGCTCTAGTCTTTTCTTAAGTCTGTGCAGTAATAGAGG 430
Dy 129614 TTTGAGGTTGCTAATCCCTTTAAAATCTAGTCTTTTTCATAGTCTGAATCTGATATAAT 129673

Qy 431 TATCGTCA--TTTATGTGACATAAAGATGGAAGGGCTTCATCTCATGTAGTAGTGGGA 488
Dy 129674 TATTATTAACTGATACTAAAATGGGTATTAATAGAGTTTTCATCATGTCTCAGTGACCCCT 129733

Qy 489 AATAGGAAGTGTGAGTGAATCTTTTAATGATGT----TTCTTTTATGAATAATTTT 544
Dy 129734 ACAGTAAACAGAGAGAGTGATCAACACGCTATCTGCTTCCTTTATGCAACTACTCTG 129793

Qy 545 AAAAGATTTGTCAGCCCTGCATGATTTTATGATGAATCATTTTGTGGTCTGTAGTTACTT 604
Dy 129794 ATGAGAGTTTGGGTACTGTGTGATCCATGATGATGCTGTTTAAATGGTATGTTGTTT 129853

Qy 605 TTAGAGATGAAGACATTTAGTGTCTCAGGAAAGCAACATTCAGATGAATGCCAATA 664
Dy 129854 CTAGAGAATAAAGTTTTTATATGCTCAGGAAAGCAATACTCGGGATGAAATTAATG 129913

Qy 665 GAGAGGTAAATTTATTTGGCATGTACATTTTCGACGCTGAGCTGTGTACATGTGTAC 724
Dy 129914 TGTATTTTGCATTTTGGCAGATTGGATGCTCAGAAATTTAGAAATCAACAATTTTC 129973

Qy 725 ACATTCTGAACATGTGTATATTGAAAT 754
Dy 129974 AAAATATTTTCTTTTAAATATTATTCAGATT 130003

```

RESULT 10

AC136055/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-123L10, *** SEQUENCING IN PROGRESS
 ***, 62 unordered pieces.

ACCESSION AC136055

VERSION AC136055.1 GI:24417926

KEYWORDS HTG; HTGS PHASE1.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 191540)
 Muzny D, Marie, Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alibrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Ayvagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 62 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Georgegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensu, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhiney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Munitasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G.,
 Olarpunsaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smajic, D., Sneed, A., Sodestren, E.,
 Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecsyk, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 191540)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (29-OCT-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1135: contig of 1135 bp in length
1136: gap of unknown length
1236: contig of 1449 bp in length
2685: gap of unknown length
2785: contig of 1303 bp in length
4088: gap of unknown length
4188: contig of 1414 bp in length
5602: gap of unknown length
5702: contig of 1483 bp in length
7185: gap of unknown length
7285: contig of 1305 bp in length
7885: gap of unknown length
8590: contig of 1514 bp in length
8690: gap of unknown length
10203: gap of unknown length
10304: contig of 1604 bp in length
11908: gap of unknown length
12008: contig of 1025 bp in length
13033: gap of unknown length
13133: contig of 1560 bp in length
14693: gap of unknown length
14793: contig of 1205 bp in length
15998: gap of unknown length
16098: contig of 1070 bp in length
17168: gap of unknown length
17268: contig of 1423 bp in length
18691: gap of unknown length
18791: contig of 1862 bp in length
20653: gap of unknown length
20752: contig of 1496 bp in length
22248: gap of unknown length
22349: contig of 1652 bp in length
24001: gap of unknown length
24101: contig of 1284 bp in length
25385: gap of unknown length
25485: contig of 1313 bp in length
26798: gap of unknown length
26898: contig of 1492 bp in length
28390: gap of unknown length
28490: contig of 1323 bp in length
29913: gap of unknown length
31236: contig of 1324 bp in length
31337: gap of unknown length
31337: contig of 1780 bp in length
33117: gap of unknown length
33217: contig of 1112 bp in length
34329: gap of unknown length
34429: contig of 2583 bp in length
37012: gap of unknown length
37112: contig of 2649 bp in length
39761: gap of unknown length
39861: contig of 1457 bp in length
41318: gap of unknown length
41418: contig of 2600 bp in length
44018: gap of unknown length
44118: contig of 1604 bp in length
45721: gap of unknown length
45821: contig of 1387 bp in length
47208: gap of unknown length
47308: contig of 1853 bp in length
49161: gap of unknown length
49261: contig of 3289 bp in length
52551: gap of unknown length
52651: contig of 2103 bp in length
54754: gap of unknown length
54854: contig of 1631 bp in length
56585: gap of unknown length
56855: contig of 1680 bp in length
58265: gap of unknown length
58365: contig of 3076 bp in length
61441: gap of unknown length

* 61541 63616: contig of 2076 bp in length
* 63617 63716: gap of unknown length
* 63717 65642: contig of 1926 bp in length
* 65643 65742: gap of unknown length
* 65743 68360: contig of 2618 bp in length
* 68361 68460: gap of unknown length
* 68461 70926: contig of 2466 bp in length
* 70927 71026: gap of unknown length
* 71027 74092: contig of 3066 bp in length
* 74093 74192: gap of unknown length
* 74193 76455: contig of 2263 bp in length
* 76456 76555: gap of unknown length
* 76556 79002: contig of 2447 bp in length
* 79003 79102: gap of unknown length
* 79103 82482: contig of 3380 bp in length
* 82483 82582: gap of unknown length
* 82583 85503: contig of 2921 bp in length
* 85504 85603: gap of unknown length
* 85604 87769: contig of 2166 bp in length
* 87770 87869: gap of unknown length
* 87870 90671: contig of 2802 bp in length
* 90672 90771: gap of unknown length
* 90772 94059: contig of 3288 bp in length
* 94060 94159: gap of unknown length
* 94160 98678: contig of 4519 bp in length
* 98679 98778: gap of unknown length
* 98779 104079: contig of 5301 bp in length
* 104080 104179: gap of unknown length
* 104180 108047: contig of 3868 bp in length
* 108048 108147: gap of unknown length
* 108148 112386: contig of 4239 bp in length
* 112387 112486: gap of unknown length
* 112487 116184: contig of 3698 bp in length
* 116185 116284: gap of unknown length
* 116285 122755: contig of 6471 bp in length
* 122756 122855: gap of unknown length

Query Match 11.4%; Score 88.6; DB 2; Length 191540;

Best Local Similarity 60.7%; Pred. No. 1e-09;

Matches 184; Conservative 0; Mismatches 109; Indels 10; Gaps 2;

Qy 371 TTTTCAGGGTTATTAACTCTTTTAAAGTCTAGTCTTTCTTAAAGTCTGTCAGTAAATAGAGG 430
Db 13980 TTTGAGCTTGCTAAATCCCTTTAAAGATCTAGGTTTCTTAAAGTCTGTCAGTAAATAGAGG 13921
Qy 431 TATCGTCATTCATGTCACATAAAAG-----ATGGAAGGGGCTTCATTTCATGTTAGTGA 484
Db 13920 TATTATTAAATTGATATTAAATAATGGACATAATAAAGACCTTTGTCATCATGTCAGTGA 13861
Qy 485 TGGAAATAGGAAAGTAGTGAAGTGAATTTTAAATAGA----TGTTTCTTTTATGAAATTAAT 540
Db 13860 CCCTAGTGGTAAACAAAGAGAGAGTGAATCAATACCTATCTGCTCTCTTTTATGCAACCAT 13801
Qy 541 TTTTAAAGATTGTCAGCCCTGTCATGATTATATGATGAATCATTTTGGTCTGTTAGTT 600
Db 13800 TCTGAGGAGAGTTGGGTGTCGATGATCCAAACATGATGATCATTTAGTGGTGTGTTGGTG 13741
Qy 601 ACTTTTAGAGAAATAGAAAGCATTTAGGCTCAGGGAAGCAACATTCAGAAATGAAATCC 660
Db 13740 ATTTCTAGAGAAATAGAAAGCTTTACATGCTCGGAGAGCAAAATCTGTTGGTGAATTT 13681
Qy 661 AAT 663
Db 13680 AAT 13678

RESULT 11

AC091418/c

LOCUS

DEFINITION

AC091418

AC091418

AC091418.5

AC091418.5

AC091418.5

AC091418.5

AC091418.5

AC091418.5

AC091418.5

245724 bp DNA linear HTG 13-NOV-2002
Rattus norvegicus clone CH230-2G6, *** SEQUENCING IN PROGRESS ***
GI:24941354
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
1 (bases 1 to 245724)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Albrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbaria, J., Benton, J., Bmaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Huly, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karleson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, C., Liu, J., Liu, W., Louissegh, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwomu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sison, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 245724)
Worley, K.C.

Direct Submission
Submitted (20-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
3 (bases 1 to 245724)
Worley, K.C.

Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Nov 13, 2002 this sequence version replaced gi:22855703.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUOY
Center clone name: CH230-2G6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 220652 bases at least Q40
Consensus quality: 223895 bases at least Q30
Consensus quality: 225929 bases at least Q20
Estimated insert size: 229375; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 245724: contig of 245724 bp in length.

FEATURES
Location/Qualifiers
1. 245724
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-2G6"
1. 1001
/notes="wgs contig"
243338..245724
/notes="wgs_contig"
misc_feature
misc_feature
ORIGIN
Query Match 11.4%; Score 88.6; DB 2; Length 245724;
Best Local Similarity 60.7%; Pred. No. 9.8e-10;
Matches 184; Conservative 0; Mismatches 109; Indels 10; Gaps 2;
Qy 371 TTTCAGGGTTATTAATCCTTTTAAAGGCTAGTTCTTCTTAAGTCTGTCAGTAATAGAGG 430
Db 52722 TTTCAGGGTTCTAATCCCTTTAAAGATCTAGGTTTCTTAAGTCTGAATGGTAATGTAAT 52663
Qy 431 TATCGTCATTCATGTCACATAAAG-----ATGGAAGGGGCTTCATTTCATGTTAGTGA 484
Db 52662 TATTTATTAATTGATATTAATAAATGACATAATAAAGACTTTTGCATCATGTCAGTGA 52603
Qy 485 TGGAAATAGGAAGTAGTGAAGTGAATTTAATAGA-----TGTTCTTTTATGAAATAAT 540
Db 52602 CCCTAGTGGTAAACACAGAGAAGTGAATCAATACCTATCTCTCTCTTTTATGCAACCAT 52543
Qy 541 TTTTAAAGATTGTCACAGCCCTGCATGATTTATGATGAATCATTTCTGGTCTCTTAGTT 600
Db 52542 TCTGAGAGAGTTTGGGTGCTGCATGATCCAAACATGAGTCAITTTAGTGGTGTGGTG 52483
Qy 601 ACTTTTAGAGATAGAAAGCATTTGTAGGCTCAGGGAAGCAAAACATTCAGAAATCAATCC 660
Db 52482 ATTCTAGAGATATAAAGCTTTTACATGCTCGGAGAAGCAAAATACCTGTGGATGAAAT 52423
Qy 661 AAT 663
Db 52422 AAT 52420
RESULT 12
AC120668/c AC120668 250169 bp DNA linear HTG 23-NOV-2002
LOCUS Rattus norvegicus clone CH230-24K10, *** SEQUENCING IN PROGRESS
DEFINITION

Best Local Similarity 60.7%; Pred. No. 9.8e-10;
Matches 184; Conservative 0; Mismatches 109; Indels 10; Gaps 2;

Qy 371 TTTCCAGGTTATTAATCCTTTTAAGGCTAGTCTTTTCTTAAGTCTGCGAGTATAGAGG 430
Db 191916 TTTGAGCTGCTAATCCCTTTTAAAGATCTAGGTTTCTTAAGTCTGAATGTAATGAAT 191857
Qy 431 TATCGTCATTCATGTGACATATAAG-----ATCGAAGGGCTTCATTCATGTAGTGA 484
Db 191856 TATTATTAAATTTGATATTAAAAATGGACATAATAAAGACTTTTGCATCATGTCAGTGA 191797
Qy 485 TGGAAATAGGAAGTAGTGAAGTGAATTTTAATAGA-----TGTTTCTTTTATGAATAAT 540
Db 191796 CCCTAGTGGTAAAAACAAGAGAAGTGAATCAATACCTATCTGTCTCTTTTATGCAACCAT 191737
Qy 541 TTTTAAAGATGTCAGCCCTGCGATGATTTATCATGAATCATTTTGTGGTCTGTTAGTT 600
Db 191736 TCTGAGGAGAGTTTGGGTGCTGATGATCCAAATGATGATCATTTAGTGGTGTGTTGGTG 191677
Qy 601 ACTTTTGAAGATAGAAGCAATTTAGGCTCAGGGAAGCAACATTCAGAAATGAAATCC 660
Db 191676 ATTTCTAGAGATATAAAGCTTTACATGCTCGGAGAAGCAATACTGTGATGAATTT 191617
Qy 661 AAT 663
Db 191616 AAT 191614

RESULT 13
AC142242 154195 bp DNA linear HTG 07-AUG-2003
LOCUS Atelerix albiventris clone LB4-81H3, WORKING DRAFT SEQUENCE, 5
DEFINITION ordered pieces.
ACCESSION AC142242
VERSION AC142242.2 GI:33469169
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Atelerix albiventris (Middle-African hedgehog)
ORGANISM Atelerix albiventris
Mammalia; Eutheria; Insectivora; Erinaceidae; Erinaceinae;
Atelerix.

REFERENCE 1 (bases 1 to 154195)
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154195)
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2003) Genome Sciences, Lawrence Berkeley National
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
REFERENCE 3 (bases 1 to 154195)
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2003) Genome Sciences, Lawrence Berkeley National
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT On Aug 7, 2003 this sequence version replaced gi:29294044.

Sequence Produced by Berkeley PGA

Web site: <http://pga.lbl.gov>

Center Code: PGABERK

Center Project Name: E015

Bac Clone Name: LB4-81H3

This sequence has been compared to sequences of other species
using VISTA (<http://www-gsd.lbl.gov/VISTA>). The results can be
viewed at:

http://pga.lbl.gov/cgi-bin/search_cwvcd?cvalue=n&value=PPARG

The order-orientation of the draft sequence was accomplished by
using:

Avid (<http://baboon.math.berkeley.edu/mavid/>),
Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

If the Bac Library Name is LB1 to LB4, please see website
for the description: <http://www-gsd.lbl.gov/cheng/BAC.html>
These libraries are available through the BACPAC Resources Center:
<http://www.chori.org/bacpac/libraryres.htm> as LBNU-1 to LBNU-4.

Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 33275: contig of 33275 bp in length

* 33276 33375: gap of unknown length

* 33376 54251: contig of 20876 bp in length

* 54252 54351: gap of unknown length

* 54352 65733: contig of 11382 bp in length

* 65734 65833: gap of unknown length

* 65834 127515: contig of 61682 bp in length

* 127516 127615: gap of unknown length

* 127616 154195: contig of 26580 bp in length.

FEATURES
source

1. 154195

/organism="Atelerix albiventris"

/mol_type="genomic DNA"

/db_xref="taxon:9368"

/clone="LB4-81H3"

ORIGIN

Query Match 8.8%; Score 68.6; DB 2; Length 154195;
Best Local Similarity 62.3%; Pred. No. 2.9e-05;
Matches 124; Conservative 0; Mismatches 74; Indels 1; Gaps 1;
Qy 558 GCCTCGCATGATTTATGATGAATCATTTTGTGGTCTCTTAGTCTACTTTTACAGAAATAGAA 617
Db 17762 GTCTGTGAATTCACCCAGGATCACTGCTAGTCTGTCTTCTCTGGGAAACAA 17821
Qy 618 AGCATTTAGGCTCAGGAAAGCAACATTCAGAAATCAATCCAAATAGAGAGGTAATTT 677
Db 17822 AGCATGATAAACACAGGGGACAGAG-CACCTCAGTGTGAAGTACAGCAGAGAAACAAGT 17880
Qy 678 TATTGGGCATGTACATTTTGGCAGCCTAGCTGTGTACATGTGTACACATTTCTGACAT 737
Db 17881 TAGTTGTGCTTGAACATTTTGGCAGCCTTGGTGTCTTGAAGACTCTGGAATGTTAAAAATG 17940
Qy 738 GTGTGTATATTGAAAAATCT 756
Db 17941 TTGTTCTTTTACITTT 17959

RESULT 14

BX465190

LOCUS

DEFINITION

complete sequence.

ACCESSION

BX465190

VERSION

BX465190.4

KEYWORDS

HTG.

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Osteichthyes;

BX465190 183103 bp DNA linear VRT 13-NOV-2003

LOCUS Zebrafish DNA sequence from clone DKEY-121A11 in linkage group 22,

complete sequence.

ACCESSION

BX465190

VERSION

BX465190.4

KEYWORDS

HTG.

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Osteichthyes;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 183103)
Barlow, K.
Direct Submission
Submitted (13-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 13, 2003 this sequence version replaced gi:32567584.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zf1sh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'drr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_erio/fishmask.shtml DXEY-121A11
is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

FEATURES
source
1..183103
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clones="DXEY-121A11"
/clone_lib="DanioKey"

ORIGIN

Query Match 6.8%; Score 53; DB 5; Length 183103;
Best Local Similarity 48.5%; Pred. No. 0.081;
Matches 176; Conservative 0; Mismatches 185; Indels 2; Gaps 1;
Qy 321 ATATTTTAAAGGATCGTTTTCGCATGTATAAATTTTCAACATTAACATTTTCAGGGTT 380
Db 159530 ATATGTTTAAAGCTACTATTACCTGACGTGTATACATTTTCAATATTTTACACACATTTG 159589
Qy 381 ATTAATCCTTTTAAAGTCTAGTGTCTTTTCTTAAGTCTGTGCGAGTAATAGAGGT--ATCGTCA 438
Db 159590 TTTTAAACCTAAAAGTTCTCTATACTATTGTGTTTTTGGCCCTGATTACAGTACATCATAT 159649
Qy 439 TTTATGTGACATAAAGATGGAAGGGCTTCATTCATGTTAGTGATGGAATAGGAAG 498
Db 159650 TTTACTAGTATTTTGAAGAATACTAGTATTTACGCTTAAAGTGGAAATTTGAATGTTAAT 159709
Qy 499 TAGGTGAAGTCATTTTAAATAGATGTTCTTTTATGAATAATTTTTTAAAGATGTCAG 558
Db 159710 TAGGTGAAGTCATTTGACACACACTGGTCTGTTGTGCAACCATCGAATAAATTTGAT 159769

Qy 559 CCTGCGCATGTTTATGATGAAATCATTTTGTGGTCTGTGTAGTACTTTTAGAGATAGAAA 618
Db 159770 AATATTGACCATATTGACCTGGAAAATTTATTTTAAAAAATTTTAAATTACAGTCAGCTTAAA 159829
Qy 619 GCATTGTAGCTCAGGAGAGACCAACATTCAGATGAAATCCCAATAGAGAGGTAAATTT 678
Db 159830 AGAATAAGACTGAATAGAAAATAGACAGAGAAAACATTAAATACCGTAAATTC 159889
Qy 679 ATT 681
Db 159890 CTT 159892

RESULT 15

CR847935

LOCUS

DEFINITION

CR847935

VERSION

CR847935.2

KEYWORDS

HTG; HTGS PHASE1

SOURCE

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 145160)

Sims, S.

Direct Submission

Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Oct 11, 2004 this sequence version replaced gi:54019745.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zf1sh-help@sanger.ac.uk

----- Project Information

Center project name: zC194N10

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 141804 bases at least Q40

Consensus quality: 142048 bases at least Q30

Consensus quality: 142375 bases at least Q20

Insert size: 144560; sum-of-contigs

Insert size: 150827; 3.5% error; agarose-fp

Quality coverage: 11.77x in Q20 bases; sum-of-contigs Quality

coverage: 11.28x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 47977: contig of 47977 bp in length

* 47978 48077: gap of 100 bp

* 48078 61854: contig of 13777 bp in length

* 61855 61954: gap of 100 bp

* 61955 114748: contig of 52794 bp in length

* 114749 114848: gap of 100 bp

* 114849 119330: contig of 4482 bp in length

* 119331 119430: gap of 100 bp

* 119431 127887: contig of 8457 bp in length

* 127888 127988: gap of 100 bp

* 127989 134017: contig of 6030 bp in length

* 134018 134117: gap of 100 bp

* 134118 145160: contig of 11043 bp in length.

* Location/Qualifiers

FEATURES

```
source
1. .145160
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-194N10"
/clone_lib="CHORI-211"
1. .47977
/note="assembly fragment:00776"
fragment_chain:1
clone_end:T7
vector_side:left"
48078. .61854
/note="assembly fragment:00558"
fragment_chain:1"
61955. .114748
/note="assembly fragment:01633"
fragment_chain:2"
114849. .119330
/note="assembly fragment:00004"
fragment_chain:2"
119431. .127887
/note="assembly fragment:00176"
fragment_chain:2"
127988. .134017
/note="assembly fragment:00071"
fragment_chain:2"
134118. .145160
/note="assembly fragment:00347"
fragment_chain:2
clone_end:SP6
vector_side:right"

ORIGIN
Query Match 6.8%; Score 52.8; DB 2; Length 145160;
Best Local Similarity 44.9%; Pred. No. 0.093;
Matches 201; Conservative 0; Mismatches 247; Indels 0; Gaps 0;
Qy 320 TATATTTAAAGGATCGTTTTGGCCATGTATAAAATTTTCAAAACATTAACCTTCAGGGT 379
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 9856 TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 9915
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 380 TATTAATCCTTTTAAAGTCTAGTTTTCTTAAAGTCTGTGCAGTAATAGAGGTATCGTCAT 439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 9916 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAG 9975
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 440 TCATGTGACATAAAGATGGAAGGGCTTCATTCACTGTAGTGTGATGGAATAGGAAGT 499
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 9976 TATTATTATTAGTATTATTATCATTTATTTATTTATTTATTTATTTATTTATTTATTT 10035
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 500 AGGTGAAGTGATTTTAAATAGATGTTCTTTTATGAAATAATTTTAAAGATTCGCCAGC 559
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 10036 ACTGTGCTGCTGTTATTTAAATATTTATTTATTTATTTATTTATTTATTTATTTATTT 10095
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 560 CCTGCATGATTTATGATGAATCATTTTGTGCTCTGTTAGTTACTTTTGTAGAAATAGAAAG 619
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 10096 ATTATTATTATTATTATAAATGTTTATTGCTTTGTACGAACGTTTAAAGTGAAAGTC 10155
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 620 CATTTAGGCTCAGGAAAGCAACATTCAGAATGAAATCCAATGAGAGGTAATTTA 679
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 10156 TTTCAATTATGCTGTATTTAGCTATTTTACCTTATATAAATAAACAATAAATAA 10215
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 680 TTTCGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCCTGAACATGT 739
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 10216 TAAAACTTTAAATATGTTAAGAAATAGACCCCTTCTCTGCACAGCAAGTCAAAAATAAA 10275
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 740 GTGTATATTGAAATCTTGTCTCTTTT 767
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 10276 TTGTATATTGTCAAAATGTCTTTTTT 10303
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:46:48 ; Search time 629.342 Seconds
(without alignments)
7308.644 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144

Perfect score: 777

Sequence: 1 tcattgtagtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004as.*
13: Geneseq2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	777	100.0	1463	2	AA19034 Human PPA
2	764.2	98.4	158417	13	ADS36461
3	383	49.3	468	2	AA19065 Human PPA
4	200.6	25.8	201	13	ADS39355
5	200.6	25.8	201	13	ADS39165 Human aut
6	196.6	25.3	201	13	ADS39169 Human aut
7	195.8	25.2	201	13	ADS39163 Human aut
8	50	6.4	5771	6	ABN80072 Human che
9	49.2	6.3	18683	6	ABL32312 Human imm
10	49.2	6.3	18683	6	ABL54333 Chemical
11	49	6.3	11422	6	ABK39937 Human che
12	49	6.3	11422	6	ABL32219 Human imm
13	48.4	6.2	13123	6	ABK31423 Signal tr
14	48.4	6.2	13123	6	ABL54364 Chemical
15	47	6.0	6880	6	ABK31322 Signal tr
16	47	6.0	6880	6	ABL70293 Chemical
17	47	6.0	6880	6	ABL70293 Chemical
18	46.8	6.0	6179	4	AA546343 Tumour su
19	46.8	6.0	6179	4	ABK31250 Signal tr
20	46.6	6.0	33053	6	ABQ67006 Human ang

c	21	46.4	6.0	50000	6	ABL56202	AmEPV gen
	22	45.2	5.8	2000	6	ABZ17494	Arabidops
	23	45.2	5.8	5718	4	AA546464	Tumour su
	24	45.2	5.8	5718	6	ABL33373	Human imm
	25	44.4	5.7	5886	6	ABL34213	Human imm
	26	44.4	5.7	5992	6	AA561208	Human gen
	27	44.4	5.7	16811	6	ABL33946	Human imm
	28	44.4	5.7	18624	6	ABL33702	Human imm
	29	44.2	5.7	5937	6	ABL34542	Human met
	30	44.2	5.7	5937	6	ABN80150	Human che
	31	44.2	5.7	5937	7	ADS99803	Human che
	32	44.2	5.7	6057	6	ABK31397	Signal tr
	33	44.2	5.7	6057	6	ABL70362	Chemical
	34	44.2	5.7	6823	6	ABL33145	Human imm
	35	44	5.7	13584	6	ABL32615	Human imm
	36	43.8	5.6	17918	6	AA561418	Human gen
	37	43.6	5.6	69727	10	ACF65374	Photorehab
	38	43.6	5.6	110000	10	ACF67367_35	Continuation (36 o
c	39	43.4	5.6	2000	6	ABZ16087	Arabidops
c	40	43.4	5.6	6849	6	ABL92253	Chemical
	41	43.4	5.6	10543	6	ABK31249	Signal tr
	42	43.4	5.6	10543	6	ABL70206	Chemical
	43	43.4	5.6	10543	6	AA561161	Human gen
	44	43.4	5.6	34548	6	ABL70603	Chemical
	45	43.2	5.6	6056	6	ABL33026	Human imm

ALIGNMENTS

RESULT 1

AA19034
ID AA19034 standard; DNA; 1463 BP.

XX AA19034;

AC AA19034;

DT 13-MAY-1999 (first entry)

XX Human PPAR-gamma-3 proximal promoter, exon A2 and intron A2.

XX Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
XX regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
XX lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
XX insulin resistance; non-insulin-dependent diabetes mellitus;
XX polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
XX inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.

XX Homo sapiens.

XX WO9905161-A1.

XX 04-FEB-1999.

XX 24-JUL-1998; 98WO-US015411.

XX 25-JUL-1997; 97US-0053692P.

XX (LIGA-) LIGAND PHARM INC.

XX (INSP) INST PASTEUR.

XX Briggs MR, Saladin RS, Auwerx J, Fajas L;

XX WPI; 1999-142844/12.

XX Newly isolated nucleic acid comprising a control region of a human
XX peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
XX identifying modulators that are useful in treating diseases associated
XX with abnormal levels of human PPAR-gamma gene expression.
XX Claim 11; Page 88; 102pp; English.

XX The present invention describes an isolated, purified or enriched nucleic
XX acid comprising a control region of a human peroxisome proliferator

activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful for screening for agents capable of modulating the expression of a human PPAR-gamma gene. These agents (modulators) form pharmaceutical compositions that are useful for treating diseases associated with high/low levels of human PPAR-gamma gene expression. The diseases include obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas, abnormalities associated with anti-human immunodeficiency virus (HIV) treatment, insulin resistance, non-insulin-dependent diabetes mellitus (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI) tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis and bowel cancer. The nucleic acids are useful for studying the role of the PPAR-gamma gene in various diseases and disorders. The structure of PPAR-gamma enables genetic studies of PPAR-gamma mutations in humans, and evaluation of its role in disorders like insulin resistance, NIDDM, and diseases associated with altered adipose tissue function, like obesity and lipodystrophic syndromes. The nucleic acids are also useful for gene therapy and the production of transgenic animals, which are useful in screening assays. The control regions of the nucleic acids enable screening for modulators of the human PPAR-gamma gene, which are useful in designing drugs for treating disorders or diseases associated with the level of PPAR-gamma gene expression. The present sequence represents human PPAR-gamma-3 proximal promoter, exon A2 and intron A2

Seq Sequence 1463 BP; 461 A; 228 C; 284 G; 490 T; 0 U; 0 Other;

Query Match 100.0%; Score 777; DB 2; Length 1463;
Best Local Similarity 100.0%; Pred. No. 2.4e-183;
Matches 777; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGTAGTGAAGTCTGTGTAAGATGTCGGGTCTCGATTTGGCGCTATTCAAGCCCTGA 60
Db TCATGTAGTGAAGTCTGTGTAAGATGTCGGGTCTCGATTTGGCGCTATTCAAGCCCTGA 427

Qy 61 TGATAAGGCTTTTGGCATTTAGATGCTGTTTGTCTTCATGGAATAACAGCTATTCAGG 120
Db TGATAAGGCTTTTGGCATTTAGATGCTGTTTGTCTTCATGGAATAACAGCTATTCAGG 487

Qy 121 ATCTCTGAGCCCTTTCAAGAGATAAGGTTGTGAATCTTAAGACCTAGGACCAATTTTACT 180
Db ATCTCTGAGCCCTTTCAAGAGATAAGGTTGTGAATCTTAAGACCTAGGACCAATTTACT 547

Qy 181 TAGATGATCTGCTCTCTGTTGTCCTCTGAAAGTCTGCTTCGTGAGGGGTGCTGCA 240
Db TAGATGATCTGCTCTCTGTTGTCCTCTGAAAGTCTGCTTCGTGAGGGGTGCTGCA 607

Qy 241 TTTGCCCTGCTAGTGTGGTGGCACACACTGCTACCTAGCTAGCTTAATAACCAT 300
Db TTTGCCCTGCTAGTGTGGTGGCACACACTGCTACCTAGCTAGCTTAATAACCAT 667

Qy 301 GTGTCATCTAGAATGAAGTTATATTTTAAAGGATCGTTTTGCCATGTATAAATTTTC 360
Db GTGTCATCTAGAATGAAGTTATATTTTAAAGGATCGTTTTGCCATGTATAAATTTTC 727

Qy 361 AAACATTAACTTTCAGGGTTAATATCTTTTAAGGTCTAGTTTTCTTAAAGTCTGTGCA 420
Db AAACATTAACTTTCAGGGTTAATATCTTTTAAGGTCTAGTTTTCTTAAAGTCTGTGCA 787

Qy 421 GTATAGAGGTPATCGTCACTTCATGACATAAAGATGGAAGGGCTTCATTCATCTTA 480
Db GTATAGAGGTPATCGTCACTTCATGACATAAAGATGGAAGGGCTTCATTCATCTTA 847

Qy 481 GTGATGGAATAGGAAGTAGGTGAAGTGAATTTTAATAGATGTTTCTTTTATGAATAAT 540
Db GTGATGGAATAGGAAGTAGGTGAAGTGAATTTTAATAGATGTTTCTTTTATGAATAAT 907

Qy 541 TTTTAAAGATGTGCCAGCCCTGCATGATTTATGATGAATCATTTTGTGCTCTGTAGTT 600
Db TTTTAAAGATGTGCCAGCCCTGCATGATTTATGATGAATCATTTTGTGCTCTGTAGTT 967

Qy 601 ACTTTTGAAGATAGGAAGCATTTGAGCTCAGGGAAGCAAACTTCAGAAATCAATCC 660
Db ACTTTTGAAGATAGGAAGCATTTGAGCTCAGGGAAGCAAACTTCAGAAATCAATCC 1027

661 AATAGAGAGGTAATAATTTATTTGGCATGTACATTTTGGGAGCCCTAGGCTGTGTACATGT 720
Db AATAGAGAGGTAATAATTTATTTGGCATGTACATTTTGGGAGCCCTAGGCTGTGTACATGT 1087

Qy 721 GTACACATTCGAAACATGCTGTATATATGAAATCTTTGCTCTCTTTTATTTGTTAAG 777
Db GTACACATTCGAAACATGCTGTATATATGAAATCTTTGCTCTCTTTTATTTGTTAAG 1144

RESULT 2
ADS36461
ID ADS36461 standard; DNA; 158417 BP.
XX AC ADS36461;
XX DT 16-DEC-2004 (first entry)
XX Human autoimmune disease-related genomic DNA sequence - SEQ ID 1675.
XX single nucleotide polymorphism detection; SNP detection;
XX rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
XX systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
XX thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
XX glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
XX primary systemic vasculitis; ds.
XX Homo sapiens.
XX OS
XX WO2004083403-A2.
XX PD 30-SEP-2004.
XX PF 18-MAR-2004; 2004WO-US008461.
XX PR 18-MAR-2003; 2003US-0455444P.
XX PR 25-APR-2003; 2003US-0465241P.
XX (APPL-) APPLERA CORP.
XX Cargill M, Begovich AB, Alexander HC;
XX WPI; 2004-728480/71.
DR New isolated nucleic acid molecule comprises at least 8 contiguous
XX nucleotides where one of the nucleotides is a single nucleotide
PT polymorphism (SNP), useful for diagnosing or treating autoimmune
PT diseases, e.g. rheumatoid arthritis.
XX Claim 16; SEQ ID NO 1675; 123pp; English.
XX The invention comprises amino acid and coding sequences containing
XX genetic polymorphisms associated with an altered risk of developing an
XX autoimmune disease (e.g. rheumatoid arthritis). The invention further
XX comprises a method of identifying an individual that has an altered risk
XX of developing an autoimmune disease, comprising detecting a single
XX nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
XX and protein sequences of the invention are useful for diagnosing and
XX treating autoimmune diseases, such as: rheumatoid arthritis, type 1
XX diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
XX bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
XX anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
XX myocarditis, Sjogren's disease, or primary systemic vasculitis. The
XX present nucleic acid represents a human autoimmune disease-related
XX genomic DNA sequence of the invention. NOTE: The present sequence is not
XX shown in the specification, but has been retrieved from the WIPO website.
SQ Sequence 158417 BP; 47887 A; 30186 C; 31475 G; 48577 T; 0 U; 292 Other;

Query Match 98.4%; Score 764.2; DB 13; Length 158417;
Best Local Similarity 98.6%; Pred. No. 1.3e-179;
Matches 766; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TCATGTAGTGAAGTCTGTAGAAATGTCGGGTCTCGATGTGGCGCTATTCAAGCCCTGA 60


```
Db 300 GGCATGTACATTTTGGCAGCCTAGCTGTGTACATGTGTACACATTTCTGAACATGTGTG 359
Qy 743 TATATTGAAATCTTGTCTCTTTTATTGTTAAG 777
Db 360 TATATTGAAATCTTGTCTCTTTTATTGTTAAG 394

RESULT 4
ID ADS39355 standard; DNA; 201 BP.
XX ADS39355;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human autoimmune disease-related SNP context sequence - SEQ ID 4569.
XX
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
XX
OS Homo sapiens.
XX
PN WO2004083403-A2.
XX
PD 30-SEP-2004.
XX
PF 18-MAR-2004; 2004WO-US008461.
XX
PR 18-MAR-2003; 2003US-0455444P.
XX
PR 25-APR-2003; 2003US-0465241P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cargill M, Begovich AB, Alexander HC;
XX
WPI; 2004-728480/71.
XX
New isolated nucleic acid molecule comprises at least 8 contiguous
nucleotides where one of the nucleotides is a single nucleotide
polymorphism (SNP), useful for diagnosing or treating autoimmune
diseases, e.g. rheumatoid arthritis.
XX
Claim 16; SEQ ID NO 4569; 123pp; English.
XX
The invention comprises amino acid and coding sequences containing
genetic polymorphisms associated with an altered risk of developing an
autoimmune disease (e.g. rheumatoid arthritis). The invention further
comprises a method of identifying an individual that has an altered risk
of developing an autoimmune disease, comprising detecting a single
nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
and protein sequences of the invention are useful for diagnosing and
treating autoimmune diseases, such as: rheumatoid arthritis, type 1
diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
myocarditis, Sjogren's disease, or primary systemic vasculitis. The
present DNA sequence represents a human autoimmune disease-related
genomic-based SNP context sequence of the invention. NOTE: The present
sequence is not shown in the specification, but has been retrieved from
the WIPO website.
XX
SQ Sequence 201 BP; 49 A; 43 C; 45 G; 63 T; 0 U; 1 Other;
Query Match 25.8%; Score 200.6; DB 13; Length 201;
Best Local Similarity 99.5%; Pred. No. 4.4e-40;
Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 102 AAAATACAGCTATTCTTAGGATCCTTTCATTAAGATGAAGTTGTGAATCCTAA 161
|||||
```

```
Db 1 AAAATACAGCTATTCTTAGGATCCTTTCATTAAGATGAAGTTGTGAATCCTAA 60
Qy 162 GACCCTAGGACCACTTTACTTTAGATGATCTGCTCTCTGGTTTCGCTCTGAAAAGTCTGCT 221
|||||
Db 61 GACCCTAGGACCACTTTACTTTAGATGATCTGCTCTCTGGTTTCGCTCTGAAAAGTCTGCT 120
|||||
Qy 222 TCGTGAGGGGTGTGCTGCATTTGCCCTTTCCTTAAGTGGTGGCACACAACTGTACTGTCA 281
|||||
Db 121 TCGTGAGGGGTGTGCTGCATTTGCCCTTTCCTTAAGTGGTGGCACACAACTGTACTGTCA 180
|||||
Qy 282 CCTTAGGCTTAATAACCATGT 302
|||||
Db 181 CCTTAGGCTTAATAACCATGT 201
|||||

RESULT 5
ADS39165
ID ADS39165 standard; DNA; 201 BP.
XX ADS39165;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human autoimmune disease-related SNP context sequence - SEQ ID 4379.
XX
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
XX
OS Homo sapiens.
XX
PN WO2004083403-A2.
XX
PD 30-SEP-2004.
XX
PF 18-MAR-2004; 2004WO-US008461.
XX
PR 18-MAR-2003; 2003US-0455444P.
XX
PR 25-APR-2003; 2003US-0465241P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cargill M, Begovich AB, Alexander HC;
XX
WPI; 2004-728480/71.
XX
New isolated nucleic acid molecule comprises at least 8 contiguous
nucleotides where one of the nucleotides is a single nucleotide
polymorphism (SNP), useful for diagnosing or treating autoimmune
diseases, e.g. rheumatoid arthritis.
XX
Claim 16; SEQ ID NO 4379; 123pp; English.
XX
The invention comprises amino acid and coding sequences containing
genetic polymorphisms associated with an altered risk of developing an
autoimmune disease (e.g. rheumatoid arthritis). The invention further
comprises a method of identifying an individual that has an altered risk
of developing an autoimmune disease, comprising detecting a single
nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
and protein sequences of the invention are useful for diagnosing and
treating autoimmune diseases, such as: rheumatoid arthritis, type 1
diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
myocarditis, Sjogren's disease, or primary systemic vasculitis. The
present DNA sequence represents a human autoimmune disease-related
genomic-based SNP context sequence of the invention. NOTE: The present
sequence is not shown in the specification, but has been retrieved from
the WIPO website.
XX
```

SQ Sequence 201 BP; 48 A; 41 C; 46 G; 65 T; 0 U; 1 Other;
Query Match 25.8%; Score 200.6; DB 13; Length 201;
Best Local Similarity 99.5%; Pred. No. 4.4e-40;
Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 125 TTGAGCCTTTCATAGAGATAGGTTGTGAATCCTAAGACCTAGGACCATTTACTTAGA 184
DB 1 TTGAGCCTTTCATAGAGATAGGTTGTGAATCCTAAGACCTAGGACCATTTACTTAGA 60
QY 185 TGATCTGCTCTCTGCTCTCTGAAAGTCTGTCGAGGGGTGCTGCATTG 244
DB 61 TGAATCTGCTCTCTGCTCTCTGAAAGTCTGTCGAGGGGTGCTGCATTG 120
QY 245 CTTTGCTTAAGTGGTGGGCACACAACTGTACTGTACCTTAGGCTTAATAACCATGTGT 304
DB 121 CTTTGCTTAAGTGGTGGGCACACAACTGTACTGTACCTTAGGCTTAATAACCATGTGT 180
QY 305 CATCTAGAAATGAAGTTATTT 325
DB 181 CATCTAGAAATGAAGTTATTT 201
RESULT 6
ADS39169
ID ADS39169 standard; DNA; 201 BP.
AC ADS39169;
DT 16-DEC-2004 (first entry)
DE Human autoimmune disease-related SNP context sequence - SEQ ID 4383.
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
OS Homo sapiens.
XX
XX
XX WO2004083403-A2.
XX
XX
XX 30-SEP-2004.
XX
XX 18-MAR-2004; 2004WO-US008461.
XX
XX 18-MAR-2003; 2003US-0455444P.
XX 25-APR-2003; 2003US-0465241P.
XX
XX (APPL-) APPLERA CORP.
XX
XX Cargill M, Begovich AB, Alexander HC;
XX
XX WPI; 2004-728480/71.
XX
XX New isolated nucleic acid molecule comprises at least 8 contiguous
XX nucleotides where one of the nucleotides is a single nucleotide
XX polymorphism (SNP), useful for diagnosing or treating autoimmune
XX diseases, e.g. rheumatoid arthritis.
XX
XX Claim 16; SEQ ID NO 4383; 123pp; English.
XX
XX The invention comprises amino acid and coding sequences containing
XX genetic polymorphisms associated with an altered risk of developing an
XX autoimmune disease (e.g. rheumatoid arthritis). The invention further
XX comprises a method of identifying an individual that has an altered risk
XX of developing an autoimmune disease, comprising detecting a single
XX nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
XX and protein sequences of the invention are useful for diagnosing and
XX treating autoimmune diseases, such as: rheumatoid arthritis, type 1
XX diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory

CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC present DNA sequence represents a human autoimmune disease-related
CC genomic-based SNP context sequence of the invention. NOTE: The present
CC sequence is not shown in the specification, but has been retrieved from
CC the WIPO website.
XX
SQ Sequence 201 BP; 50 A; 35 C; 47 G; 68 T; 0 U; 1 Other;
Query Match 25.3%; Score 196.6; DB 13; Length 201;
Best Local Similarity 99.5%; Pred. No. 4.3e-39;
Matches 196; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCATGTAGTAAAGTCTGTGTAGAAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
DB 5 TCATGTAGTAAAGTCTGTGTAGAAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 64
QY 61 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTCTCATGGAATAACAGCTATTCTAGG 120
DB 65 TGATAAGGCTTTTGGCATTAGATGCTGTTTGTCTTCTCATGGAATAACAGCTATTCTAGG 124
QY 121 ATCCTTGAGCCTTTTCATAAGAGATAAGTGTGTGAATCCTTAGACCTTAGGACCATTTACT 180
DB 125 ATCCTTGAGCCTTTTCATAAGAGATAAGTGTGTGAATCCTTAGACCTTAGGACCATTTACT 184
QY 181 TAGATGATCTGCTCTCT 197
DB 185 TAGATGATCTGCTCTCT 201
RESULT 7
ADS39163
ID ADS39163 standard; DNA; 201 BP.
XX
XX ADS39163;
XX
XX 16-DEC-2004 (first entry)
XX
XX Human autoimmune disease-related SNP context sequence - SEQ ID 4377.
XX
XX single nucleotide polymorphism detection; SNP detection;
XX rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
XX systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
XX thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
XX glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
XX primary systemic vasculitis; ds.
XX
XX Homo sapiens.
XX
XX WO2004083403-A2.
XX
XX 30-SEP-2004.
XX
XX 18-MAR-2004; 2004WO-US008461.
XX
XX 18-MAR-2003; 2003US-0455444P.
XX 25-APR-2003; 2003US-0465241P.
XX
XX (APPL-) APPLERA CORP.
XX
XX Cargill M, Begovich AB, Alexander HC;
XX
XX WPI; 2004-728480/71.
XX
XX New isolated nucleic acid molecule comprises at least 8 contiguous
XX nucleotides where one of the nucleotides is a single nucleotide
XX polymorphism (SNP), useful for diagnosing or treating autoimmune
XX diseases, e.g. rheumatoid arthritis.
XX
XX Claim 16; SEQ ID NO 4377; 123pp; English.
XX
XX The invention comprises amino acid and coding sequences containing

XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP007537.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX Claim 1; SEQ ID NO 285; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 18683 BP; 5439 A; 130 C; 3594 G; 9520 T; 0 U; 0 Other;
XX
XX Query Match 6.3%; Score 49.2; DB 6; Length 18683;
XX Best Local Similarity 46.5%; Pred. No. 0.072;
XX Matches 228; Conservative 0; Mismatches 258; Indels 4; Gaps 2;
XX
QY 290 TTAATAACCATGTGTCATCTAGAGTGAAGTTATATTTTAAAGAGATCGTTTTGCCATG 349
Db |||||
6355 TTAAAAATAAGTCGTGGTTGATATTTATAATAATTAATAATGATAGTATTATTATT 6414
QY 350 TATAAATTTTCA--ACATTAACCTTCAGGGTTATTAATCCCTTTTAAAGTCTAGTTTTC 407
Db |||||
6415 TTAATTTTGGAGGAATGTATTTTGTATGTGTTTAAGTATATATTATTTTGTGGA 6474
QY 408 TTAAGTCTGTGCAGTAATAGAGTATCGTCATTCATGTGACATAAAGATGGAAGGGC 467
Db |||||
6475 ATTAAAAATGTGGTTTTCATAGTTTATTTTAAATGATTTTGTATTATTATTATT 6534
QY 468 TTCAATTCATGTTAGTGATGGAATAGGAAGTAGGTGAAGTGAATTTTAATAGATGTTTCT 527
Db |||||
6535 ATTGTAGTTTATTATTAGTGATTATTAATAATAAATAAATAATTTATAGTAATTTT 6594
QY 528 TTATGAATAATATTTTAAAGATGTCAGGCCCTCGATGATTTATGATGATCAATTTTG 587
Db |||||
6595 TTTAAAAAATATTTATATATGATAGTAAATTTATGATTTTATGATTTTAAATTTTAA 6654
QY 588 TGGTCTGTAGTTACTTTTAGAGATAGAAAGCATTTGAGGCTCAGGGAAGC--AAACA 645
Db |||||
6655 TATTAATTAATATAGTTGGTTATTATTTTATATTTAGTTGAAGGGATTTTAGAGA 6714
QY 646 TTCAGATGAATCCAAATAGAGAAGGTAAATTTATTTGGGCATGATTTTGGCAGCCT 705
Db |||||
6715 ATGTTTTTAAAAATATGTTTAAAGTTTAAAGTTTGTGAGAGAGTGGTTTAGGTTGTA 6774
QY 706 AGGCTGTACATGTCACACATTCGACATGTCGTATATTGAAATCTGTCCTTT 765
Db |||||
6775 TGAATTTTAAATTTTGTATTATTAATTTAATTTTTCGTAATATTTTAAATATTGTTTG 6834
QY 766 TTATTTGTTA 775
Db |||||
6835 TTTTGTATTA 6844

RESULT 10
ABL54333
ID ABL54333 standard; DNA; 18683 BP.
XX AC ABL54333;
XX 29-JUL-2002 (first entry)
XX Chemically treated apoptosis gene #17.
XX Apoptosis; HIV; Bloom syndrome; cardiopathy; neurodegenerative disorder;
XX Herpes simplex virus; renal ischaemia; amyotrophic lateral sclerosis;
XX cancer; ds.
XX Unidentified.
XX WO200177164-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-EP003969.
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017444/02.
XX Chemically modified sequences of genes associated with apoptosis are
XX useful to determine methylation patterns of genomic DNA samples for
XX diagnosis of associated diseases such as cancer.
XX Claim 1; Seq ID #33; 24pp; English.
XX This invention relates to chemically pre-treated DNA of genes associated
XX with apoptosis. The nucleic acids are used to allocate patients for
XX specific therapy for HIV infection, Bloom syndrome, cardiopathy, aging,
XX neurodegenerative disorders, Herpes simplex virus infection, renal
XX ischaemia, amyotrophic lateral sclerosis, solid tumours and cancers. This
XX nucleotide sequence represents a chemically treated apoptosis gene. Even
XX SEQ ID numbers are the complementary DNA strands to the odd SEQ ID
XX numbers. The sequence data for this patent is not represented in the
XX printed specification but is based on information supplied by the
XX European patent office
XX Sequence 18683 BP; 5439 A; 130 C; 3594 G; 9520 T; 0 U; 0 Other;
XX
XX Query Match 6.3%; Score 49.2; DB 6; Length 18683;
XX Best Local Similarity 46.5%; Pred. No. 0.072;
XX Matches 228; Conservative 0; Mismatches 258; Indels 4; Gaps 2;
XX
QY 290 TTAATAACCATGTGTCATCTAGAGTGAAGTTATATTTTAAAGAGATCGTTTTGCCATG 349
Db |||||
6355 TTAAAAATAAGTCGTGGTTGATATTTATAATAATTAATAATGATAGTATTATTATT 6414
QY 350 TATAAATTTTCA--ACATTAACCTTCAGGGTTATTAATCCCTTTTAAAGTCTAGTTTTC 407
Db |||||
6415 TTAATTTTGGAGGAATGTATTTTGTATGTGTTTAAGTATATATTATTTTGTGGA 6474
QY 408 TTAAGTCTGTGCAGTAATAGAGTATCGTCATTCATGTGACATAAAGATGGAAGGGC 467
Db |||||
6475 ATTAAAAATGTGGTTTTCATAGTTTATTTTAAATGATTTTGTATTATTATTATT 6534
QY 468 TTCAATTCATGTTAGTGATGGAATAGGAAGTAGGTGAAGTGAATTTTAATAGATGTTTCT 527
Db |||||
6835 TTTTGTATTA 6844


```
XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP007537.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX Claim 1; SEQ ID NO 192; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX PS Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;
XX
XX Query Match 6.3%; Score 49; DB 6; Length 11422;
XX Best Local Similarity 46.4%; Pred. No. 0.071;
XX Matches 235; Conservative 0; Mismatches 265; Indels 7; Gaps 2;
Qy 272 TGTACTGTCACCTTGGCTTAATTAACCATGTCATCTAGAGTGAAGTATATATTTTAAAA 331
Db 6619 TTTTATTTTAAATTTTAAAAATTTTAAAAATTTTAAAAATTTTAAAAATTTTAAAAATTTT 6678
Qy 332 AGGATCGTTTGGCCATGATATATTTTCAACATTAACCTTCAGGGTATTAATCTTT 391
Db 6679 ATGTTGTTGTTAAAGATTAAGACGATATATATTTTAAAGTAGTGTATTA-----TT 6733
Qy 392 TAAGGTCAGTTTCTTAAAGTCGTGCAGTAATAGAGGTATCGTCAATCATGTGACATA 451
Db 6734 GAAATTTATATAGTTTAAATTTGTTGTTAATATATATATAGTAGAGAGATAGTATATAGA 6793
Qy 452 AAAGATGGAAGGGGCTTCATTCATGTTAGTGTGATGGAATAGGAAGTAGGTGAAGTCAT 511
Db 6794 AAATAAGAAATATAATGTTTATAGATATAAGTATGTAATTTAAAAATTTAGTTTATTTAA 6853
Qy 512 TTTAATAGATGTTTCTTTTATGAATATATTTTAA--AAGATTGTCAGCCCTCATGAT 569
Db 6854 TTGAATGAGATGTTTTCATATGATATTTTAAAAATTTTATATTTTAAATTTTATTTAA 6913
Qy 570 TTATGATGAATCATTTTGTGTCGTGTTAGTGTACCTTTTAGAGAAATAGAAAGCAATGTAGGC 629
Db 6914 GTTATATAAATTTTAGTATTTGTTAATTTTATTTAGAGATTTTGTAGGTATATATA 6973
Qy 630 TCAGGGAAGCAACATTCAGAAATCAATCAATAGAGAGGTTAAATTTATTTGGGCATG 689
Db 6974 TGATGAGAAGTTATAAATTTTGGTTAAGATAGGAAAAATAAAAATGAATATTTTATTA 7033
Qy 690 TACATTTTGGCAGCTAGGCTGTGATCATGTGTACACATCTCGAACATGTGTGATATTG 749
Db 7034 AATTTTAAAGAGATATATTTTGTAAATTTTGTAGTGTATATGTTTGGAGATATATTT 7093
Qy 750 AAAATCTGTCTCTTTTATTTTATTTGTTAA 776
Db 7094 ATGTTTGGGGTTTATTTTATTTTGGAA 7120
```

```
RESULT 13
ABK31423
ID ABK31423 standard; DNA; 13123 BP.
XX AC ABK31423;
XX DT 23-APR-2002 (first entry)
XX DE Signal transduction associated gene modified complementary DNA #133.
XX KW Human; signal transduction associated gene; cytosine methylation state;
XX KW CpG island; signal transduction associated disease; solid tumour; cancer;
XX KW antitumour; cytostatic; mutant; ds.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200200926-A2.
XX PD 03-JAN-2002.
XX PF 29-JUN-2001; 2001WO-EP007472.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-147896/19.
XX PT Oligonucleotide for diagnosis and therapy of diseases associated with
XX PT signal transduction e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with signal transduction.
XX PS Claim 1; SEQ ID NO 266; 24pp; English.
XX CC The present invention relates to chemically modified DNA sequences of
XX CC signal transduction associated genes. The DNA sequences are chemically
XX CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
XX CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
XX CC the cytosine methylation state (CpG islands) of these genes, and a method
XX CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
XX CC genes associated with signal transduction. The genomic DNA can be
XX CC obtained from cells or cellular components which contain DNA. e.g. cell
XX CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
XX CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
XX CC brain, heart, prostate, lung, breast or liver, histologic object slides,
XX CC and all their possible combinations. The sequences of the invention are
XX CC useful for the diagnosis and therapy of diseases associated with signal
XX CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
XX CC chemically pretreated genomic DNA sequences of different genes associated
XX CC with signal transduction, or their complementary sequences. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office
XX SQ Sequence 13123 BP; 3999 A; 255 C; 2547 G; 6320 T; 0 U; 2 Other;
XX
XX Query Match 6.2%; Score 48.4; DB 6; Length 13123;
XX Best Local Similarity 49.1%; Pred. No. 0.1;
XX Matches 157; Conservative 0; Mismatches 161; Indels 2; Gaps 1;
Qy 284 TTAGGCTTAATTAACCATGTCATCTAGATGAGTTATATTTAAAAAGATCGTTT 343
Db 2473 TTATGTAATAAAATTTTATTTTATTTTATTTAGTTTAAATTTTAAAAAGATTTTGT 2532
Qy 344 GCCATGTATAAAATTTTCAACATTAACCTT--TCAGGGTATTAATTCCTTTTAAGGTCAG 401
Db 2533 TTTTGTTTTAAATTTATTTATGTTATTTTATTTAGTTTATGAGGTTTATTTTATTTGATTT 2592
```


Job time : 631.342 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 15:16:42 ; Search time 197.983 Seconds
(without alignments)
6421.701 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144

Perfect score: 777

Sequence: 1 tcattgtagtaagactgtgt.....gtctcttttttattgttaag 777

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgm2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgm2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgm2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgm2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgm2_6/ptodata/1/ina/PTCUS_COMB.seq.*
- 6: /cgm2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.4	5.8	1743	4	US-09-248-796A-11015
C 2	44.4	5.7	54576	4	US-09-949-016-15954
C 3	44.4	5.7	54576	4	US-09-949-016-15955
C 4	41.8	5.4	231129	4	US-09-949-016-16110
C 5	41.8	5.4	266293	4	US-09-949-016-11934
C 6	41.4	5.3	601	4	US-09-949-016-156535
C 7	40.4	5.2	1141	4	US-09-806-708B-22
C 8	40.4	5.2	360470	4	US-09-949-016-13173
C 9	39.8	5.1	175236	4	US-09-949-016-14353
C 10	39.8	5.1	298336	4	US-09-949-016-16600
C 11	39.6	5.0	1141	4	US-09-806-708B-22
C 12	39	5.0	601	4	US-09-949-016-142482
C 13	39	5.0	181429	4	US-09-949-016-12372
C 14	39	5.0	181430	4	US-09-949-016-15772
C 15	38.4	4.9	1783	4	US-09-679-409-9
C 16	38.4	4.9	1898	4	US-09-679-409-8
C 17	38.4	4.9	1923	4	US-09-679-409-12
C 18	38.4	4.9	1981	4	US-09-679-409-6
C 19	38.4	4.9	1992	4	US-09-679-409-7
C 20	38.4	4.9	2075	4	US-09-679-409-11
C 21	38.4	4.9	2132	4	US-09-679-409-5
C 22	38.4	4.9	2215	4	US-09-679-409-4
C 23	38.4	4.9	2226	4	US-09-679-409-2
C 24	38.4	4.9	2308	4	US-09-679-409-13
C 25	38.4	4.9	2309	4	US-09-679-409-10
C 26	38.4	4.9	2431	4	US-09-679-409-3
C 27	38.4	4.9	2612	4	US-09-679-409-15

ALIGNMENTS

RESULT 1

US-09-248-796A-11015/c
; Sequence 11015, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstein et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA

TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.132

CURRENT APPLICATION NUMBER: US/09/248.796A

CURRENT FILING DATE: 1999-02-12

PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR FILING DATE: 1998-02-13

PRIOR APPLICATION NUMBER: US 60/096,409

PRIOR FILING DATE: 1998-08-13

NUMBER OF SEQ ID NOS: 28208

SEQ ID NO 11015

LENGTH: 1743

TYPE: DNA

ORGANISM: Candida albicans

US-09-248-796A-11015

Query Match

Best Local Similarity 5.8%; Score 45.4; DB 4; Length 1743;

Matches 124; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy	492	AGGAAAGTAGGTGAAGTGAATTTTAAATAGATGTTCTTTTATGAATAATATTTTAAAGAT	551
Db	1431	AGTTAAATATTTAAATTTTGTAAATAATTGAAATTTTCAATTAATATTAATAAAAAAT	1372
Qy	552	TGTCACCCCTGCATGATTTATGATGATCATCTTTTGGTCTGTTAGTACTTTTAGAGA	611
Db	1371	TTTGAATAATGATTTAATAATTTTCGTTAATTAATGATGATTTTATGATA	1312
Qy	612	ATAGAAAGCATTTAGGCTCAGGAAAGCAACATTCAGATGAATCCCAATAGAGAAGG	671
Db	1311	TAATAATGCAATAATTTGGTAATATATTTGAATAATTTTCAGATTAATCGTTCACCGAAC	1252
Qy	672	TAAATTTATTTGGCATGTACATTTTGGCAGCTAGGCTGTGTACATGTGACATTTCT	731
Db	1251	TAAATAACCTTGATAAAATTTTTTTTTTAAATACATAATTTCAATAAAATTTAGTCTTAATCCG	1192
Qy	732	GAACATGTGTGTATA	746
Db	1191	TGAGGTGGTTGAGA	1177

RESULT 2

US-09-949-016-15954
; Sequence 15954, Application US/09949016

QY 465 GCCTTCATTCAGTTAGTGAAGAAATAGGAAGTAGTGAAGTGAATTTTAATAGATGTT 524
Db 72756 TATATATGAATATATATTTATGTATTATATATTTATATATATATATATATATTT 72815
QY 525 TCCTTTATGAATAATTTTAAACATTTGTCAGCCCTGCATGATTTATGATGAATCAAT 584
Db 72816 ATATACCTTAATTT 72875
QY 585 TTGTGGTCTGTTAGTACTTTTGA 609
Db 72876 ATATATTTTATATATAATATATATA 72900

RESULT 5
US-09-949-016-11934
; Sequence 11934, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11934
; LENGTH: 266293
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11934

Query Match 5.4%; Score 41.8; DB 4; Length 266293;
Best Local Similarity 45.5%; Pred. No. 1.3;
Matches 148; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 285 TAGGCTTAATAACCATGTCATCTAGAAATGAAGTATATTTTAAAAAGGATCGTTTTTG 344
Db 59740 TATATTTTATATATAAAATATATATTTTATTTATATATAATATATATATTTT 59799
QY 345 CCATGTATAAAATTTTCAACATTAACCTTTCAGGGTTATTAATCCCTTTAAGTCTAGTTT 404
Db 59800 TTATTTATATAAATATATATTTATATATATATATATATATATATATATATAT 59859
QY 405 TTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGCACATAAAAGATGAAAGG 464
Db 59860 TATATATATATATATATATATATATATATATATATATTTTATGATTTATATATAT 59919
QY 465 GCCTTCATTCAGTTAGTGAAGAAATAGGAAGTAGTGAAGTGAATTTTAATAGATGTT 524
Db 59920 TATATATGAATATATTTTATGATTTATATATATATATATATATATATATATAT 59979
QY 525 TCCTTTATGAATAATTTTAAACATTTGTCAGCCCTGCATGATTTATGATGAATCAAT 584
Db 59980 ATATACCTTAAT 60039
QY 585 TTGTGGTCTGTTAGTACTTTTGA 609
Db 60040 ATATATTTTATATATAATATATATA 60064

RESULT 6
US-09-949-016-156535/c
; Sequence 156535, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

Query Match 5.2%; Score 40.4; DB 4; Length 1141;

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156535
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-156535

Query Match 5.3%; Score 41.4; DB 4; Length 601;
Best Local Similarity 45.2%; Pred. No. 0.23;
Matches 147; Conservative 1; Mismatches 177; Indels 0; Gaps 0;

QY 285 TAGGCTTAATAACCATGTCATCTAGAAATGAAGTATATTTTAAAAAGGATCGTTTTTG 344
Db 438 TATATTTTATATATAAAATATATATTTTATTTATATATAATATATATATTTT 379
QY 345 CCATGTATAAAATTTTCAACATTAACCTTTCAGGGTTATTAATCCCTTTAAGTCTAGTTT 404
Db 378 TTATTTATATATAATATATATTTTATATATATATATATATATATATATATAT 319
QY 405 TTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGCACATAAAAGATGAAAGG 464
Db 318 TATATATATTTATATATATATATAATATATATATATATATATATATATATATAT 259
QY 465 GCCTTCATTCAGTTAGTGAAGAAATAGGAAGTAGTGAAGTGAATTTTAATAGATGTT 524
Db 258 TATATATGAATATATATATTTTATGATTTATATATATATATATATATATATATAT 199
QY 525 TCCTTTATGAATAATTTTAAAAAGATTTGCCAGCCCTGCATGATTTATGATGAATCAAT 584
Db 198 ATATACCTTAAT 139
QY 585 TTGTGGTCTGTTAGTACTTTTGA 609
Db 138 ATATATTTTATATATAATATATATA 114

RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match


```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16600
; LENGTH: 298336
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(298336)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16600

```

Query Match	5.1%	Score 39.8;	DB 4;	Length 298336;
Best Local Similarity	47.7%	Pred. No. 4.7;		
Matches 116;	Conservative 0;	Mismatches 127;	Indels 0;	Gaps 0;
Qy	528	TTTATGAAATAAATTTTAAAGATTGTCACGCCCTGCATGATTATATGATGAATCATTTTG	587	
Db	273892	TATATATATATACACCGTATAGAATATATAGTTATATATATATATACAGTATAGAATATA	273833	
Qy	588	TGGTCGTAGTTACTTTTCAGAGATAGAAAGCATTTGTAGCTCAGGGAAGCAACATT	647	
Db	273832	TAGTTAGTTATATACAGTATAGAATATATAGTTATATATACAGTATAGAATATATAGTTA	273773	
Qy	648	CAGAAATGAAATPCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTTCGACGCTAG	707	
Db	273772	TATATTTATATACAGTATAGAATATATAGTTATATATTTATATACAGTATAGAATATA	273713	
Qy	708	GCTGTGTACATGCTACACATTTCTGAACATGTGTGTATATTTGAAAATCTTGTCTCTTTTT	767	
Db	273712	GTTATATACAGTATAGAATATATAGTTATATATACAGTATAGAATATATAGTTATATATA	273653	
Qy	768	TAT	770	
Db	273652	TAT	273650	

```

RESULT 11
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: Consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

```

	Query Match	5.1%	Score 39.6;	DB 4;	Length 1141;	
	Best Local Similarity	8.8%;	Pred. No. 0.85;			
	Matches	55;	Conservative 252;	Mismatches 320;	Indels 0;	Gaps 0
Qy	145	AAGTTGTGAATCTTAAGACCCCTAGGACCATTACTTAGATGATCTGCTCTCGTTCGT	204			
	:	: : : :	:	: : : :	:	: : : :
Db	77	RWKGWKYKWBVCANNITSBRYHARRWDKMKTAYBMTMTNKWGKTGWHRHYRWRAWBDT	136			
Qy	205	CCTCTGAAACGTCTGCTCGTGAGGGGTGCTGCATTTTCCTTCCTAAGTGGTGTGGC	264			
	:	: : : : : :	:	: : : : : :	:	: : : : : :
Db	137	VDRHYVTANNNNAWTTCMMMDKDDRRTRWWKKNNNATGWDGDDTKTHMNNNGCBVTVMWV	196			
Qy	265	ACACAAGTGTACTGTCACTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATAT	324			

197	RYKTRDWSBKRMNYGMBWKNQWSYDVITYYWWDDCMCKRVWRVTRGRMRTMVAWB	256
325	TTTAAAAAGGATCGTTTTTGGCCATGTATAAATTTTCAAACATTAACTTTTCAGGGTTATTA	384
257	TAHRRYYNNGWTBAWAYRRWTMNNNNNAKAMCKRAKYGNRAVNSTCTTTWKSSTTKV	316
385	ATCCTTTTAAGGCTAGTTTTCCTTAAGCTGTGCAGTAAATAGAGGTATCGTCATTCATG	444
317	RTSCWANNCRAGDANDHKHMKWKSAAAGVYWNNNNNNNWTTKYCARHBAWDWVHSAWKK	376
445	TGACATAAAAGATCGAAAGGGGCTTCATTCATGTTTAGTGATGGAATAAGGAAGTAGGTG	504
377	WHANAHSRKWTBYKRTKTMVNNNGTTTWMKRWMAWTWKMDMDWBGTYNNNNNGRTYY	436
505	AAGTGAATTTTAATAGATGTTTCTTTTATGAATAAATTTTAAAAAGATTGTCAGCGCCTCG	564
437	GWTKGKSWTYYYKWKANCKWRADWDHKTCENNTTWKTKTYWNNCYKWSMTNGKSHRBA	496
565	ATGATTATGAATGAATCATTTTGTGGTCTGTAGTTACTTTTAGAAGATAGAAAGCATTG	624
497	AAVYTMWMMWRRYAHANNNNNDWYKACTWYKYBVCWSKWNNTYAAWTTKSSWNTSRYY	556
625	TAGCTCAGGGAAGCAAAACATTACGAATGAAATCCAATAGAGAAGGTAAATTTATTTCGG	684
557	RWKTNNSWRRSDTRFSGRANNYARABHYGYKWNTRWWBSHTWBHBAGAAHYWMBMY	616
685	GCATGTACATTTTGGCAGCCCTAGGCTGTGTACATGTGTACACATCTCTCAACATGTGTGA	744
617	BAKCHCKWAWKAKYAGAGGSNNNNNNNNNNNNNNATCARDDYYAASRWYMANAKW	676
745	TATTGAAAACTCTGCTCTTTTTTATT	771
677	YYXKBAANNAYYTHANNWGCWNNATD	703

RESULT 12
US-09-949-016-142482/c
? Sequence 142482, Application US/09949016
? Patent No. 6812339
? GENERAL INFORMATION:
? APPLICANT: VENTER, J. Craig et al.
? TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
? WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
? FILE REFERENCE: CL001307
? CURRENT APPLICATION NUMBER: US/09/949.016
? CURRENT FILING DATE: 2000-04-14
? PRIOR APPLICATION NUMBER: 60/241,755
? PRIOR FILING DATE: 2000-10-20
? PRIOR APPLICATION NUMBER: 60/237,768
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: 60/231,498
? PRIOR FILING DATE: 2000-09-08
? NUMBER OF SEQ ID NOS: 207012
? SOFTWARE: FastSeq for Windows version 4.0
? SEQ ID NO 142482
? LENGTH: 601
? TYPE: DNA
? ORGANISM: Human
US-09-949-016-142482

	Query Match	5.0%;	Score 39;	DB 4;	Length 601;	
	Best Local Similarity	46.5%;	Pred. No. 1;			
	Matches 126;	Conservative 0;	Mismatches 145;	Indels 0;	Gaps 0	
Qy	325	TTTAAAGGATCGTTTTTGCCATGTATAAATTTTCAAAACATTAACTTTCAGGGTTATTA	384			
Dd	406	TTTACAACCTTTAGAATTCTTTATATAAACTATATAACTTTAATACAAAAATTTTCAAA	347			
Qy	385	ATCCTTTTAGGCTAGTTTTTTCCTAGTCGTGCAGTAATAGAGTATCGTCATTCATG	444			
Dd	346	TTTCACITACAAAATCTCTGTGGTTACTCTCTGAGAAAGGTGTGTTTTTTTTTTTTTTT	287			

Qy	445	TGACATAAAAGATGAAAGGGCTTCATCTCATGTTAGTATGGAAATAGGAAAGTAGGTG	504
Db	286	TTTGGACTAATGTACCAAAATGACCACATTTTTTGTGAAAGTGGGCCAAAGTAAAGTTTTAA	227
Qy	505	AAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTTAAAGATTGTCAGCCCTGC	564
Db	226	TTATATAGTACACCAATTGCATGTTTTGAAACATCATGTATGAATTAGTAAACTTTTT	167
Qy	565	ATGATTTAATGATGAATCATTTTGTGGTCTGT	595
Db	166	TCTTTGTTTTATAAAACCGTTTTTTTCTTT	136
RESULT 13			
US-09-949-016-12372			
; Sequence 12372, Application US/09949016			
; Patent No. 6812339			
; GENERAL INFORMATION:			
; APPLICANT: VENTER, J. Craig et al.			
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF			
; FILE REFERENCE: CLO01307			
; CURRENT APPLICATION NUMBER: US/09/949.016			
; CURRENT FILING DATE: 2000-04-14			
; PRIOR APPLICATION NUMBER: 60/241,755			
; PRIOR FILING DATE: 2000-10-20			
; PRIOR APPLICATION NUMBER: 60/237,768			
; PRIOR FILING DATE: 2000-10-03			
; PRIOR APPLICATION NUMBER: 60/231,498			
; PRIOR FILING DATE: 2000-09-08			
; NUMBER OF SEQ ID NOS: 207012			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 12372			
; LENGTH: 181429			
; TYPE: DNA			
; ORGANISM: Human			
US-09-949-016-12372			
Query Match			
Best Local Similarity 5.0%; Score 39; DB 4; Length 181429;			
Matches 126; Conservative 0; Mismatches 145; Indels 0; Gaps 0;			
Qy	325	TTTAAAGAGTATGTTTGGCCATGATATAAATTTTCAAACAATTAACTTTCAGGGTTATTA	384
Db	11497	TTTCAAACTTTTGAAGATTATCTTATATAAATCTATAAATCTATAAATTTTCAAA	11556
Qy	385	ATCCTTTTAAGTCTAGTTTTTCTTAAAGTCGTGAGTAATAGAGGTATCGTCATTCTG	444
Db	11557	TTTCACCTTACAAATCTCTGGTTACTTCTGAGAAAGGTGTGTTTTTTTTTTTTTTT	11616
Qy	445	TGACATAAAAGATGGAAGGGCTTCATTCATGTTAGTATGGAAATAGGAAAGTAGGTG	504
Db	11617	TTTGGACTAATGTACCAAAATGACCACATTTTTTGTGAAAGTGGGCCAAAGTAAAGTTTTAA	11676
Qy	505	AAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTTAAAGATTGTCAGCCCTGC	564
Db	11677	TTATATAGTACACCAATTCATGTTTTGAAACAAATCATGTATGAATTAGTAAACTTTTT	11736
Qy	565	ATGATTTATGATGAATCATTTTGTGGTCTGT	595
Db	11737	TCTTTGTTTTATAAAACCGTTTTTTTCTTT	11767
RESULT 14			
US-09-949-016-15772			
; Sequence 15772, Application US/09949016			
; Patent No. 6812339			
; GENERAL INFORMATION:			
; APPLICANT: VENTER, J. Craig et al.			
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF			
; FILE REFERENCE: CLO01307			
; CURRENT APPLICATION NUMBER: US/09/949.016			

```
; NAME/KEY: polyA signal
; LOCATION: 1757..1762
; NAME/KEY: allele
; LOCATION: 132
; OTHER INFORMATION: 8-130-143 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 132
; OTHER INFORMATION: 8-126-286 : polymorphic base A or G
US-09-679-409-9

Query Match      4.9%; Score 38.4; DB 4; Length 1783;
Best Local Similarity 45.9%; Pred. No. 2.1; 196; Indels 2; Gaps 1;
Matches 168; Conservative 0; Mismatches 0;

Qy 314 TGAAGTTATATTTTAAAGGATCGTTTTGCCATGTATAAAATTTTCAAAACATTAACTTT 373
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 374 CAGGTTATTAATCCTTTTAAAGTCTAGTTTTCTTAAAGTCTGTGCAGTAATAGAGGTAT 433
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 1475 TAGAGATGGCTAACCATCTATGAAGTGTATTGGATGAGCATGGTAAACAATATTATGAAA 1416
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 434 CGTCATTCACTGCACATAAAAGATGGAAGGGCTTCAATTCATTGTAGTGATGGAAATAG 493
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 1415 TAGAGTTTTCTCACTTAAATCCAAAGAGGCCCTGAACAGTCAATTTAGAAAAGAA 1356
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 494 GAAAGTAGGTGAAGTGAATTTTAAATAGATGTTTCTTTTATGAAATAAATTTTAAAGATTG 553
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 1355 AAAAAATACCTGAATAATATTTTCAAGCGTTAAATTTAGATATTAGTATAAATGAAAC 1296
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 554 TCCAGCCCTGCATGATTTATGATGAATCAATTTTGTGTCGTAGTTACTTTTAGAGAAAT 613
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 1295 TTCAACACATGPACATATATAAGTAAC--TTTGTGTTATTTAAGCATTACCAATAAAATTT 1238
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 614 AGAAGCAATTTAGGCTCAGGGAAGCAACATTCAGAAATCAATCCAAATAGAGAAGTA 673
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 1237 CAAATTTATTTCTTACTTTAGATTTTAGAAGAAATTAAGTAAAGAGATCAGCCAGAGTTC 1178
Db      |||      |||      |||      |||      |||      |||      |||      |||
Qy 674 AATTTA 679
Db      |||      |||      |||
Qy 1177 TGTGTA 1172
```

Search completed: November 8, 2005, 19:52:49
Job time : 199.983 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 18:06:53 ; Search time 1005.4 Seconds
(without alignments)
6391.142 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144

Perfect score: 777

Sequence: 1 tcattgtagtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
- 25: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.2	6.3	18683	16	US-10-311-455-285
2	49.2	6.3	18683	17	US-10-240-452-33
3	49	6.3	11422	16	US-10-311-455-192
4	49	6.3	11422	18	US-10-257-166-18
5	48.4	6.2	13123	17	US-10-240-452-64

c	6	48	6.2	595	13	US-09-925-065A-814225	Sequence 814225,
c	7	48	6.2	1693	14	US-10-027-632-256987	Sequence 256987,
c	8	48	6.2	1693	14	US-10-027-632-256988	Sequence 256988,
c	9	48	6.2	1693	18	US-10-027-632-256987	Sequence 256987,
c	10	48	6.2	1693	18	US-10-027-632-256988	Sequence 256988,
c	11	48	6.2	3673778	17	US-10-312-841-2	Sequence 2, Appl1
c	12	47	6.0	6880	18	US-10-221-613-183	Sequence 183, Appl
c	13	46.8	6.0	6179	19	US-10-221-714A-65	Sequence 65, Appl
c	14	46.6	6.0	33053	20	US-10-433-793-36	Sequence 36, Appl
c	15	46.6	6.0	3673778	17	US-10-312-841-1	Sequence 1, Appl1
c	16	46.4	6.0	50000	22	US-10-706-635-26	Sequence 26, Appl
c	17	45.2	5.8	2000	9	US-09-938-842A-5299	Sequence 5299, Ap
c	18	45.2	5.8	2000	11	US-09-938-842A-5299	Sequence 5299, Ap
c	19	45.2	5.8	5718	16	US-10-311-455-1346	Sequence 1346, Ap
c	20	45.2	5.8	5718	19	US-10-221-714A-186	Sequence 186, Ap
c	21	44.4	5.7	5886	16	US-10-311-455-2186	Sequence 2186, Ap
c	22	44.4	5.7	5992	18	US-10-221-613-167	Sequence 167, Ap
c	23	44.4	5.7	16811	16	US-10-311-455-1919	Sequence 1919, Ap
c	24	44.4	5.7	18624	16	US-10-311-455-1675	Sequence 1675, Ap
c	25	44.2	5.7	655	13	US-09-925-065A-776150	Sequence 776150,
c	26	44.2	5.7	657	13	US-09-925-065A-790118	Sequence 790118,
c	27	44.2	5.7	5937	16	US-10-240-485-95	Sequence 95, Appl
c	28	44.2	5.7	6823	16	US-10-311-455-1118	Sequence 1118, Ap
c	29	44	5.7	13584	16	US-10-311-455-588	Sequence 588, App
c	30	43.8	5.6	17918	18	US-10-221-613-381	Sequence 381, App
c	31	43.4	5.6	2000	9	US-09-938-842A-3892	Sequence 3892, Ap
c	32	43.4	5.6	2000	11	US-09-938-842A-3892	Sequence 3892, Ap
c	33	43.4	5.6	6849	19	US-10-240-589C-62	Sequence 62, Appl
c	34	43.4	5.6	10543	18	US-10-221-613-120	Sequence 120, App
c	35	43.2	5.6	6056	16	US-10-311-455-999	Sequence 999, App
c	36	43.2	5.6	7516	16	US-10-311-455-33	Sequence 33, Appl
c	37	43	5.5	6565	19	US-10-221-714A-188	Sequence 188, App
c	38	43	5.5	13321	19	US-10-221-714A-144	Sequence 144, App
c	39	42.8	5.5	6022	19	US-10-221-714A-383	Sequence 383, App
c	40	42.8	5.5	7072	18	US-10-221-613-347	Sequence 347, App
c	41	42.8	5.5	12393	16	US-10-311-455-1235	Sequence 1235, Ap
c	42	42.6	5.5	7676	16	US-10-240-485-152	Sequence 152, App
c	43	42.6	5.5	21537	16	US-10-311-455-1971	Sequence 1971, Ap
c	44	42.4	5.5	1032	18	US-10-282-122A-16418	Sequence 16418, A
c	45	42.4	5.5	7384	16	US-10-311-455-719	Sequence 719, App

ALIGNMENTS

RESULT 1

US-10-311-455-285
; Sequence 285, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Methylation of Cytosine
; FILE REFERENCE: 5013.1014
; CURRENT FILING DATE: 2002-12-16
; PRIOR FILING DATE: 2001-07-02
; PRIOR FILING DATE: 2000-06-30
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 285
; LENGTH: 18683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)

US-10-311-455-285

Query Match	6.3%	Score 49.2	DB 16	Length 18683
Best Local Similarity	46.5%	Pred. No. 0.45		
Matches 228	Conservative 0	Mismatches 258	Indels 4	Gaps 2
Qy	290	TTAATACCATGTCATCTAGATGAAGTATATATTTTAAAGGATCGTTTTGCCATG	349	
Db	6355	TTAAAAATAGTCGTGGTTGATTAATTTATAAATAATTAATAATGTAGATAGTTATTATTATT	6414	
Qy	350	TATAAATTTTCA--AACATTAACTTTCAGGGTTAATAATCCTTTAAAGGCTCTAGTTTTTC	407	
Db	6415	TTAAATTTTCAGGAAATGTATTTTGTAAATGTGTTAAGTATATATATTTTGTTTGGA	6474	
Qy	408	TTAAGTCGTGCAGTAATAGAGGTATCGTCAATTCATGTGACATAAAAAGATGGAAAGGGC	467	
Db	6475	ATTAATAATTCGTGGTTTGTGATAAGTTTTTTTTTTTAAATGATTTTTTGTATTATTATTATT	6534	
Qy	468	TTCAATTCATGTTAGTGTAGTGGNAATAGNAAGTAGGTGTAAGTGATTTTAAATAGATGTTCT	527	
Db	6535	ATTGTAGTGTTTTATTATGTAGTGATTAATAATAATAAATAATATTTATAGTAATATTATTT	6594	
Qy	528	TTTATGAATAAATTTTTAAAGATGTTCCAGCCCTGCATGATTTATGATGAATCATTTTG	587	
Db	6595	TTTAAAAAATATTTATTATATAGTAGTTAATTATGTATTTTTTTTGATTTAATATTTAA	6654	
Qy	588	TGGTCGTGTAGTTACTTTTATAGAAATAGAAGCAATCTGTAGCTCAGGGAAGC--AAACA	645	
Db	6655	TATTAATTAATATAGTTGTGTTATTTATTTTATATGTTAGTTGAAGGGGATATTTAGAGA	6714	
Qy	646	TTCAGATGAAATCCATAGAGAAGGTAAATTTATTTTGGCATGTACATTTTGGCAGCCT	705	
Db	6715	ATGTTTTTAAATAATATGTTTATAGGTTAAAGTTTGTTCAGAGGAGTGGTTTAGGTTGGTA	6774	
Qy	706	AGGCTGTGTACATGCTGTACACATTTCTCAACATGTGTGTATATTTGAAAATCTTGTCTCTTT	765	
Db	6775	TGATTTTAAATTTTGTATTATTAATTTTATTTTGTAAATATTTTAAATATTTGTTTG	6834	
Qy	766	TTTATTCGTTA	775	
Db	6835	TTTTTGATTA	6844	

```

RESULT 2
US-10-240-452-33
; Sequence 33, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240.452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 33
; LENGTH: 18683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-33

```

Query Match
6.3%; Score 49.2; DB 17; Length 18683;

Best Local Similarity 46.5%; Pred. No. 0.45;
Matches 228; Conservative 0; Mismatches 258; Indels 4; Gaps 2;

290 TTAATAACCATGTGTCATCTAGAAATGAAGTTATATTTTAAAAAGGATCGTTTTTGGCATG 349
6355 TTAATAATAAGTCGTGGTTTGATAAATTATAAATAAATAAATAGTAGATATTATTATT 6414

350 TATAAAATTTTCA--AACATTAACTTTTCAGGGTTATTAATCCCTTTAAGTCTAGTTTTC 407
6415 TTAATTTTTTGAGGAATTTGTAATTTTGTGAATGTGTTTAAAGTATTTATATATTTTGTGG 6474

408 TTAAGTCGTGTCAGTAATAGAGGTATCGTCATTCATGTGCACATAAAAAAGATGGAAGGGC 467
6475 ATTAATAATTTGGTTTGTAAAGTTTTTTTTTTAAATGATTTTGTATTATTATTATT 6534

468 TTCAATCATGTTAGTAGTGAATAGAAAGTAGGTAAGTGATTTTAAATAGATGTTTCT 527
6535 ATTGTAGTTTTTATTTTAGTGAATTTATATAATAAAAAATAAATTTATAGTAATATTTT 6594

528 TTTATGAAATAAATTTTTAAAAAGATTGTCACAGCCCTGCATGATTTATATGATGAATCATTTTG 587
6595 TTTAAAAAATAATTATATATATGATTAATTTATGTAATTTTTTTTGAATTAATAATTTAA 6654

588 TGGTCTGTTAGTTACTTTTATAGAAATAGAAAGCATTTGTAGGCTCAGGGAAGC--AAACA 645
6655 TATTAAATTAATATAGTTGTCTTATTATTTTATTTATTTGTAGTTGAAGGGGATATTTAGAGA 6714

646 TTCAGAAATGAATCCAAATAGAAAGGTAAATTTTATTTGGGCATGTACATTTTGGCAGCCT 705
6715 ATGTTTTTAAATAATGTTTTAGGTTAAAGTTTGTTCAGAGGAGTGGTTTAGGTTGGTA 6774

706 AGGCTGTTACATGTTGACACATCTGAAACATGTTGTGTATATTGAAAACTTGTCTCTCTT 765
6775 TGAATTTTAAATTTTGTATTAAATTTTAAATTTTGTAAATATTTTAAAAATATTTTGTTTG 6834

766 TTTATTGTTA 775
6835 TTTTGTATTA 6844

RESULT 3
US-10-311-455-192
; Sequence 192, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BEBLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System h
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 192
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-192

Query Match	6.3%	Score 49;	DB 16;	Length 11422;
Best Local Similarity	46.4%	Pred. No. 0.41;		
Matches 235;	Conservative	0;	Mismatches 265;	Indels 7;
				Gaps 2;

QY 272 TGTACTGTCACCTTAGGCTTAATAACCATGTGTCACTAGAAATGAAGTTATATTTTAAAA 331

Query Match 6.2%; Score 48; DB 18; Length 1693;
Best Local Similarity 46.2%; Pred. No. 0.31;
Matches 159; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Query Match	6.2%;	Score 48;	DB 18;	Length 1693;
Best Local Similarity	46.2%;	Pred. No. 0.31;		
Matches 159;	Conservative 0;	Mismatches 185;	Indels 0;	Gaps 0;
Qy	411	AGTCCTGCAGTAATAGAGGTATCGTCATTTCAATGTGCACATAAAAGATGGAAGGGGCTTC	470	
Db	475	AGCAGTGGAAATTTTGAGGAAGTGTGTTCGCAATCTACAGTGAAATGGACTATTTCTAG	416	
Qy	471	ATTTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGAATTTTAAATAGATGTTTCTTTTT	530	
Db	415	AGCGAAGAAGTATATTGGGATAAATGTTAGTGAGTGGTAAAGTGAGCACAGAGTCTCTATT	356	
Qy	531	ATGAAATAATTTTTAAAAAGATTTGTCAGCGCTGCATGATTTATGATGAATCATTTTGTGG	590	
Db	355	ATTAAGTAATTTTTTGGCAAGACTTTATCAGGGGCAAAATGTCAGATTGCAATGGGTATTTCTGA	296	
Qy	591	TCCTGTTAGTTACTTTTTAGAGAATPAGAAAGCATTTGTAGGCTCAGGGAAAGCAACATTCAG	650	
Db	295	ACTGAAAGTGAAAGAGTTAAACTGGGAGTCATAGTCTAAATATCTATTTTAAATATATTA	236	
Qy	651	AATGAAATCCAAATAGAGAAGGTAAATTTATTTTGGGCATGTACATTTTGGCAGCGCTAGGCT	710	
Db	235	TCTGAAATATGATAAAGGATGGATATTAGACTCCCATGTTACATGTTACTTATTTCTGT	176	
Qy	711	GTGTACATGTGTACACATTTCTGAAACATGTGTGTATATTTGAAAT	754	
Db	175	TAATAATAAGAAAGTAATCTTTAAGCATGTATGAATATTCAGACT	132	

```

RESULT 10
US-10-027-632-256988/c
; Sequence 256988, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIORITY APPLICATION NUMBER: US 60/218,006
; PRIORITY FILING DATE: 2000-07-12
; PRIORITY APPLICATION NUMBER: US 60/198,676
; PRIORITY FILING DATE: 2000-04-20
; PRIORITY APPLICATION NUMBER: US 60/193,483
; PRIORITY FILING DATE: 2000-03-29
; PRIORITY APPLICATION NUMBER: US 60/185,218
; PRIORITY FILING DATE: 2000-02-24
; PRIORITY APPLICATION NUMBER: US 60/167,363
; PRIORITY FILING DATE: 1999-11-23
; PRIORITY APPLICATION NUMBER: US 60/156,358
; PRIORITY FILING DATE: 1999-09-28
; PRIORITY APPLICATION NUMBER: US 60/146,002
; PRIORITY FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

```


US-10-221-714A-65
; Sequence 65, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 65
; LENGTH: 6179
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-65

Query Match 6.0%; Score 46.8; DB 19; Length 6179;
Best Local Similarity 53.0%; Pred. No. 1.1;
Matches 122; Conservative 0; Mismatches 107; Indels 1; Gaps 1;
QY 319 TTAATTTTAAAGGATCGTTTTGGCCATGATATAAATTTTCAACATTAATTTTCAGGG 378
DB 3255 TTAATTTTAAAGGATCGTTTTGGCCATGATATAAATTTTCAACATTAATTTTCAGGG 378
QY 379 TTAATTAATCCCTTTAAGGCTAGTTTCTTAAGTCTGTCGAGTAATAGAGTATCGTCA 438
DB 3314 ATATATATAAAGGATATATAGTGTGTTTAAATGTTGTAATGTAATGTAATTT 3373
QY 439 TTCAATGACATATAAAGATGAAAGGGCTTCATTCATGTTAGTATGATGAAATAGGAAAG 498
DB 3374 TAATAGAGGAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 3433
QY 499 TAGGTGAAGTATTTAATAGATGTTCTTTTATGAAATAATTTTAAAAA 548
DB 3434 TAGTATAGATTTTGTGTTATATTTATTTATTTTAAAGTATATTTTATATAA 3483

RESULT 14
US-10-433-793-36
; Sequence 36, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 36
; LENGTH: 33053
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-36

Query Match 6.0%; Score 46.6; DB 20; Length 33053;
Best Local Similarity 46.7%; Pred. No. 2.6;
Matches 148; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 290 TTAATAACCATGTCATCTAGAATGAGTTATATTTTAAAGGATCGTTTTTGCCTATG 349
DB 14921 TTTATGATTTATTTTATTTAGTATAAAGTATTTATAAGAGATTAGTTTGTATTATAGG 14980
QY 350 TATAAATTTTCAACATTAATCTTTTCAAGGGTTTAAATCTTTTAAAGGCTCTAGTTTCTT 409
DB 14981 AATAATATATAATGATGTTTATAATGTTTGAATGTAATTTAGTTTAAATTTTTTGT 15040
QY 410 AAGCTGTGCAAGTATAGAGTATCGTCATTCATCTGTCATATAAAGATGGAAGGGCTT 469
DB 15041 TATACGTTTTTGTGGAAGGTTGTGGTGTGTTTTTGTGATTTTATTAAGAAATGTTTTTGG 15100
QY 470 CATTCATGTTAGTATGGAATAGGAAGTAGGAGGAGTATTTTAAAGATGTTTCTTT 529
DB 15101 TTGATGTTTTAGTAAATGTTAGTGGTGTAATTTTGTGAATATATTTATATATTTTTT 15160
QY 530 TATGAAATAATTTTAAAGATTTGTCCAGCCCTGCATGATTTTATGATGAATCATTTTGTG 589
DB 15161 TATAGATATTTTGTGATATATAAAGTATTTTGTGATTTTGTGAGATTTTATTCAG 15220
QY 590 GTCTGTTAGTTACTTTT 606
DB 15221 AAAATTTTTTTTTTGT 15237

RESULT 15
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 6.0%; Score 46.6; DB 17; Length 3673778;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 148; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 307 TCTAGAATGAAGTTATATTTTAAAGGATCGTTTTTCCCATGATATAAATTTTCAACAT 366
DB 1671903 TTTATATTTATTTTGTGGAAGGAATGTAATTTAGTAAATTTTCGGTTGTATATAT 1671962
QY 367 TAACTTTCAAGGTTTATTAATCTTTTAAAGTCTAGTTTCTTAAAGTCTGTCAGTAATA 426
DB 1671963 TGGTTGTAGTTTATTAATTTTGTAGTAGTATATTTAGTTAGTAGTAGTTGTAA 1672022
QY 427 GAGGTATCGTCATTCATGTGACATATAAAGATGGAAGGGCTTCATTCATGTTAGTATG 486
DB 1672023 TTAAGATTTATTTGGTTAAGTTTGAAGTAATTTTGTATTATTTTAGTTTATTCGG 1672082
QY 487 GAAATAGGAAGTAGGAGTATTTTAAAGATGTTTCTTTTATGAATAATTTTAA 546
DB 1672083 TTTTGTGAAGATAGATAATACGGAGATAATAGGTAATTTTATGATTTATTTTATTTA 1672142
QY 547 AAGATTGTCAGCCCTGCATGATTTATGATGAATCATTTTGTGCTGTTAGTTACTTTT 606

Db 1672143 TAGTTTTTAAATGGTGGTGTATTTTTTATATAATATTGTGAATAATTTTTTAAGATTGT 1672202

Qy 607 AGAGAAATAGAAAGCATT 623

Db 1672203 TGGGACGTAATATCGTT 1672219

Search completed: November 8, 2005, 23:53:38
Job time : 1012.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 22:16:08 ; Search time 4014.41 Seconds
(without alignments)
7367.447 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144

Perfect score: 777

Sequence: 1 tcattagtagtaagactgtgt.....gtctctttttttattgttaag 777

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	101.6	13.1	376	2 BF736184	BF736184 PM4-KT000
C 2	63.4	8.2	358	9 CC774457	CC774457 CH240_95D
C 3	62.4	8.0	390	1 AU234451	AU234451 AU234451
C 4	62.4	8.0	400	1 AU278852	AU278852 AU278852
C 5	58.2	7.5	1101	9 CNS0039G	AL0633921 Drosophila
C 6	54.6	7.0	919	9 CNS005RL	AL061409 Drosophila
C 7	53.8	6.9	1101	9 CNS005EV	AL069706 Drosophila
C 8	53.8	6.9	1101	9 CNS00LT2	AL078714 Drosophila
C 9	53.8	6.9	1101	9 CNS016LI	AL106896 Drosophila
C 10	52.4	6.7	1200	9 CNS016CO	AL106578 Drosophila
C 11	52.4	6.7	1146	9 CNS021G2	AL176843 Tetraodon
C 12	51.8	6.7	1530	9 AG382019	AG382019 Mus muscu
C 13	51.6	6.6	500	1 AU088479	AU088479 AU088479
C 14	51.6	6.6	1201	9 CNS016BY	AL106552 Drosophila
C 15	51.4	6.6	945	9 CNS04DOK	AL285149 Tetraodon
C 16	50.2	6.5	1190	9 CNS02ON7	AL206908 Tetraodon
C 17	49.6	6.4	1101	9 CNS00GCK	AL072030 Drosophila
C 18	48.8	6.3	928	9 CNS00DKY	AL071865 Drosophila
C 19	48.6	6.3	996	9 CNS00FUF	AL071063 Drosophila
C 20	48.4	6.2	1001	9 CNS01400	AL103554 Drosophila
C 21	48.4	6.2	1101	9 CNS0100X	AL098379 Drosophila
C 22	48.4	6.2	1201	9 CNS0162X	AL106227 Drosophila
C 23	48.4	6.2	1101	9 CNS001FB	AL060732 Drosophila
C 24	47.8	6.2	842	8 BZ696583	BZ696583 SP_Ba008

25	47.6	6.1	979	9 CNS0161W	AL106190 Drosophila
C 26	47.6	6.1	1315	3 CR657214	CR657214 Tetraodon
C 27	47.4	6.1	813	9 BX218284	BX218284 Danio rer
C 28	47.4	6.1	1092	9 CNS020K7	AL175696 Tetraodon
C 29	47.4	6.1	1169	9 CNS06KHQ	AL402900 T3 end of
C 30	47.2	6.1	1101	9 CNS000D1	AL065414 Drosophila
C 31	47	6.0	1065	8 BZ696860	BZ696860 SP_Ba008
C 32	47	6.0	1094	9 CNS012FZ	AL101513 Drosophila
C 33	47	6.0	1101	9 CNS00265	AL097271 Drosophila
C 34	46.8	6.0	510	1 AV778356	AV778356 AV778356
C 35	46.8	6.0	609	9 CNS025K2	AL182171 Tetraodon
C 36	46.8	6.0	1031	9 CNS000CF2	AL059199 Drosophila
C 37	46.8	6.0	1101	9 CNS003BD	AL064091 Drosophila
C 38	46.8	6.0	1101	9 CNS000ES1	AL069797 Drosophila
C 39	46.6	6.0	792	9 CG116113	CG116113 PUJDS79TB
C 40	46.6	6.0	870	9 CG209274	CG209274 OGLEB38TV
C 41	46.6	6.0	878	9 CNS0187R	AL108593 Drosophila
C 42	46.6	6.0	894	8 BZ687506	BZ687506 PUBHR65TD
C 43	46.6	6.0	906	9 CG209262	CG209262 OGLEB38TH
C 44	46.4	6.0	1896	9 CG753083	CG753083 P048-1-C0
C 45	46	5.9	516	5 BM883909	BM883909 rc24b11.y

ALIGNMENTS

RESULT 1
BF736184/c
LOCUS BF736184 376 bp mRNA linear EST 10-JAN-2001
DEFINITION PM4-KT0005-151100-001-b03 KT0005 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF736184
VERSION BF736184.1 GI:12062858
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
Goldman,G.H., Carvalho,A.P., Mateukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongenseel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM4kt2-PM4-KT0005-
151100-001-b03kt3=2000-11-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 21
High quality sequence stop: 376.

FEATURES

Location/Qualifiers
1..376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev stage="Adult"
/clone lib="KT0005"
/note="Organ: bladder tumor; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products

derived from ORBSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 13.1%; Score 101.6; DB 2; Length 376;
 Best Local Similarity 95.8%; Pred. No. 8.4e-14;
 Matches 115; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 Qy 1 TCATGTAGTAACTGTGTAGAGTGGGTCTCGATGTGGCGCTATTCAAGCCCTGA 60
 Db 122 TCATGTAGTAACTGTGTAGAGTGGGTCTCGATGTGGCGCTATTCAAGCCCTGA 63
 Qy 61 TGATAAGCCTTTGGCATTAGATCTGTTTGTCTTCATGGAATAACAGCTATTCAGG 120
 Db 62 TGATAAGCCTTTTGTCTATTAGATCTGTTTGTCTTCATGG-ATACACAGCTATTCAGG 4

RESULT 2

CC774457/c
 LOCUS CH240_95D2.TV CHORI-240 Bos taurus genomic clone CH240_95D2,
 DEFINITION genomic survey sequence.
 ACCESSION CC774457
 VERSION CC774457.1 GI:32326748
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 REFERENCE 1 (bases 1 to 358)
 AUTHORS Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
 Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
 Womack,J.E., de Jong,P.J. and Lewin,H.A.
 TITLE Bovine BAC end sequences from CHORI-240 library
 JOURNAL Unpublished (2003)
 COMMENT Other_GSSs: CH240_95D2.TJ
 Contact: Harris Lewin
 Department of Animal Sciences
 University of Illinois at Urbana Champaign
 1201 W. Gregory Dr., Urbana, IL 61801, USA
 Tel: 217 333 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu

Clones are derived from the bovine BAC library CHORI-240
 (<http://www.chori.org/bacpac/bovine240.htm>). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.choi.org).
 Clones may be purchased from BACPAC Resources
 (<http://www.chori.org/bacpac/ordering-information.htm>). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBBMC) by the University of Illinois at Urbana
 Champaign, USA with funds provided by grant NO. AG202-34480-11828
 from USDA-CGREES and AG99-35205-8534 from USDA/NRI (Livestock
 Genome Sequencing Initiative)
 Plate: 95 row: D column: 2
 Seq primer: T7
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..358
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="bred: Hersford"
 /db_xref="taxon:9913"
 /clone="CH240_95D2"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="vector: pFARBAC1.3; Site 1: MboI; Site 2: MboI;
 Hersford bull L1 Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 8.2%; Score 63.4; DB 9; Length 358;
 Best Local Similarity 71.5%; Pred. No. 0.00014;
 Matches 98; Conservative 0; Mismatches 36; Indels 3; Gaps 1;
 Qy 283 CTTAGGCTTAATAACCATGTGTCATCTAGAAATGAAGTTATATTTTAAAGAGATCGTTTT 342
 Db 140 CTTTAGGTTAGCAGCATGTGCCCTCTAGAACAAAGCTTGTTATTTTCAATAATGTCATCTT 81
 Qy 343 TGCATGTATAAATTTTCAAAACAT---TAACTTTAGGGTTATTAATCCCTTTTAAGGTCT 399
 Db 80 TGCTGTGTATAAATTTTATCAATCAACCTATCTTTTGAGGTTAGTAATCCCTTTTAAGATCT 21
 Qy 400 AGTTTTTCTTAAAGTCTG 416
 Db 20 AGTTTTTCTTAAAGTCTG 4

RESULT 3

AU234451/c
 LOCUS AU234451 390 bp mRNA linear EST 21-SEP-2001
 DEFINITION AU234451 Bovine placenta cDNA Bos taurus cDNA clone Cln153 3',
 mRNA sequence.
 ACCESSION AU234451
 VERSION AU234451.1 GI:15719669
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 REFERENCE 1 (bases 1 to 390)
 AUTHORS Gohma,H., Lejukole,H.Y., Taniguchi,Y., Yamada,T., Akagi,S.,
 Yasue,H. and Sasaki,Y.
 TITLE Analysis of expressed sequence tags from a cDNA library of bovine
 placenta
 JOURNAL Unpublished (2001)
 COMMENT Contact: Takahisa Yamada
 Graduate School of Agriculture
 Kyoto University
 Sakyo-ku, Kitashirakawa, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-6323
 Fax: 81-75-753-6340
 Email: tyamada@ikans.ikans.kais.kyoto-u.ac.j
 This clone was obtained from a 3' end cDNA library.

FEATURES

source

Location/Qualifiers
 1..390
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone="Cln153"
 /tissue_type="placenta"
 /clone_lib="Bovine placenta cDNA"

ORIGIN

Query Match 8.0%; Score 62.4; DB 1; Length 390;
 Best Local Similarity 69.1%; Pred. No. 0.00025;
 Matches 123; Conservative 0; Mismatches 36; Indels 19; Gaps 2;
 Qy 600 TACTTTTAGAATAGAAAGCATGTAGGCTCAGGGAAGCAACATTCAGATGAATC 659
 Db 299 TAGTCTCTAGAAACACAAAGTGTTTAAAGCTCAGGGAAGCAACATTCAGATGAATC 240
 Qy 660 CAATAGAGAGGTAAATTTTATTTGGCATCTACATTTTGGCAGCTAGGCTGTGATCATG 719
 Db 239 CAAC-----ATGTACTTCTGCTTATTCATTTGACAGATGTTGTTCTAGT- 192
 Qy 720 TGTACACATCTCAACATGTGTGTATATGAAATCTTGTCTCTTTTATTGTTAAG 777
 Db 191 -----ATTTCGAATATGTGTATATTAATAATCTTCTCTTTTCAGTTGTCAG 141

```

RESULT 4
LOCUS      AU278852
DEFINITION AU278852 Cloned bovine placenta cDNA Bos taurus cDNA clone
ACCESSION AU278852
VERSION    AU278852.1 GI:21682162
KEYWORDS   EST.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE  1 (bases 1 to 400)
AUTHORS   Oishi,M., Yamada,T., Goma,H., Lejukole,H.Y., Taniguchi,Y. and
            Sasaki,Y.
            EST analysis of cloned bovine fetus and placenta
            Unpublished (2002)
            Contact: Masahito Oishi
            Graduate School of Agriculture
            Kyoto University
            Sakyo-ku Kitashirakawa, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-6331
            Fax: 81-75-753-6340
            Email: oishi@kais.kans.kais.kyoto-u.ac.jp.
FEATURES   Location/Qualifiers
            source
            1..400
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /clone="placental153"
            /dev_stage="60 embryonic day"
            /clone_lib="Cloned bovine placenta cDNA"

ORIGIN
Query Match      8.0%; Score 62.4; DB 1; Length 400;
Best Local Similarity 69.1%; Pred. No. 0.00025;
Matches 123; Conservative 0; Mismatches 36; Indels 19; Gaps 2;

QY      600 TACTTTAGAGAAAGCAATCTAGGCTCAGGGAAGCAAAACATTCAGAAATCAATC 659
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      92 TAGTTCTAGAGAACAAAGTGTGTTTAAAGCTCAGGGAAGCAAAACATTCAGAAATCAATC 151

QY      660 CAATAGAGAAGTAAATTTATTTGGCGATGACATTTTGGCAGCCTAGGCTGTGTACATG 719
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      152 CAAC-----ATGCTCTCTGCTTATCCATTTTGACAGATGAGTGTGCTAAGT- 199

QY      720 TGTACACATCTGACATGTGTGATATGAAATCTGTCTCTTTTATTTATGTTAAG 777
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      200 -----ATTTGAATATGTGTATATTAATAATCTTCTTTTTCAGTTGTTTCAG 250

RESULT 5
CNS0039G/c
LOCUS      CNS0039G
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL063921
VERSION     AL063921.1 GI:49411778
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
AUTHORS
TITLE      Direct Submission
JOURNAL
COMMENT    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a

```

```

collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammoss in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES   Location/Qualifiers
            1..1101
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACR08K10"
            /clone_lib="RPCI-98"
            /note="end : TET3"

ORIGIN
Query Match      7.5%; Score 58.2; DB 9; Length 1101;
Best Local Similarity 14.3%; Pred. No. 0.0031;
Matches 66; Conservative 223; Mismatches 172; Indels 0; Gaps 0;

QY      309 TAGAATGAAGTATATTTTAAAGGATCGTTTGGCCATGTATATAATTTCAACATTA 368
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1049 TWTATATWWWATWDTWWDKWWATAAKTDTATWMTATRADWAGRDGAGKDR 990

QY      369 ACTTTCAGGGTTATTAATCCTTTAAGGCTAGTGTCTTCTTAAGTCTGTCAGTAATAGA 428
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      989 DAATDADGAGRRDGGRRKDKDKDKGDDDKGKKKKAAXAAKATKWDWDWDKDK 930

QY      429 GGTATCGTCATTCATGACATAAAGATGAAAGGGCTTCATTCATGTTAGTCATGGA 488
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      929 WDGAQRKADDDGAGDKDDGKGDADDTDGTDKDDDKDDDKDKGWTGWTGATWAW 870

QY      489 AATAGGAAGTAGTGAAGTCAATTTAATAGATGTTTCTTTTATGAATAATTTTAAAA 548
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      869 AATDWWMGWADADWTTWDAADDDWADWDWAWKWDWDAWARTADRDWDGDRAGK 810

QY      549 GATTGTCAGCCCTGCGATGATTATGATGAATCAATTTGTGCTGTAGTTACTTTTAG 608
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      809 RGGARKDRKGRADDDKDDAATWTTTDTDDDKWKTDTWTWADRTWD 750

QY      609 AGAATAGAAACATTTAGGCTCAGGGAAGCAAAACATTCAGAAATGAAATCAATAGAGA 668
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      749 RDDDDDRDRAGTAGRKWRRTWKRKRDRTRWDDADADDTARDDRRRRGDDGADAGK 690

QY      669 AGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTACATGTGTACAT 728
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      689 TGRKRRDRDRATWRTDADWADAAWTTTDTDDDKRRRRKGRRRRTTAAADWD 630

QY      729 TCTGAACATGTGTATATTCGAAATCTGTCTCTTTTAA 769
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      629 WTWKAWDWKDWKTRADRWDRWAADTWDARKADRWAKA 589

RESULT 6
CNS005RL/c
LOCUS      CNS005RL
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR12F23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL061409
VERSION     AL061409.1 GI:4943512
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
1. 919
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR12F23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 7.0%; Score 54.6; DB 9; Length 919;
Best Local Similarity 22.5%; Pred. No. 0.022;
Matches 86; Conservative 128; Mismatches 168; Indels 0; Gaps 0;

QY 299 ATGTGTCATCTAGAAATGAAGTTATATTTTAAAGGATCGTTTTGCCATGTATAAATTT 358
DB 912 AWKDKKKKADAKADAKAWKDDDDAKWKKKKKKKKKKKKKKKKKKADAKAKAKAKK 853
QY 359 TCAAAACATTAACCTTCAGGGTTATTAATCCCTTTTAAAGTCTAGTTTTCTTAAGTCTGTG 418
DB 852 AKDGDWKKDAGTCAKTTGKTKGTGKWTWTTDWTWAWTTTWTATDAGWATGTG 793
QY 419 CAGTAATAGAGGTATCGTCATCTGACATTAAGATGAAGGGGCTTCATCATGT 478
DB 792 WTDGKKGKKKGAATAAKGKWTWTTWAAWTDKWTAWAAWRDAATRTWTDTAAGKKKK 733
QY 479 TAGTGATGGAATAGGAAGTAGGTGAAGTATTTAATAGATGTTCTTTTATGAATA 538
DB 732 KTKKWAATWGTAKAWATWKKKKKDTWKWKTTWTDNAWTTTWTWTWTWTWT 673
QY 539 ATTTTAAAGATGTCACAGCCCTGCATGATTTAATGATGAATCATTTTGGTCTGTAG 598
DB 672 TTTTAAAWATTTTGTGTTTTTATATGDKAAGTAWWTTTAAATTTNGTGGKGGKKG 613
QY 599 TTACTTTTAGAGATAGAAAGCATTTGAGGCTCAGGGAAGCAACATTCAGATGAAT 658
DB 612 KKGKKKGTGKKKGGKK 553
QY 659 CCAATAGAGAGGTAAATTTAT 680
DB 552 GGAATTGAAAGGWTATGTTT 531

RESULT 7

CNS00EVL/c 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

VERSION AL069706.1 GI:4949849

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 6.9%; Score 53.8; DB 9; Length 1101;
Best Local Similarity 34.1%; Pred. No. 0.035;
Matches 166; Conservative 76; Mismatches 243; Indels 2; Gaps 1;
QY 290 TTAATAACCATGTGTCATCTAGAAATGAAGTTATATTTTAAAGGATCGTTTTGCCATG 349
DB 971 WBAATWATWATWATWATTTTAAATAAATAATWAAWTTTATTTTATTTTATTTT 912
QY 350 TATAAATTTTCAACATTAACCTTCAGGGTTATTAATCCCTTTTAAAGTCTAGTTTTCTT 409
DB 911 ATWTWTTTATATWTTWAAAWAAWAAWTTTATTTTWTWTTTATWATTAATTAATTT 852
QY 410 AAGTCTGTGCAGTAATAGAGGTATCGTCATCTGACATATAAAGATGGAAGGGCTT 469
DB 851 TTWTTATTTWTTATATWATWTTTATWTTTATWTTTATWTTTATWTTTATWTTTATW 792
QY 470 CATTCTGTTAGTGTGGAATAGGAAGTAGGTAAGTGAATTTTAAATAGATGTTCTTT 529
DB 791 TATWTTATWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTW 732
QY 530 TATGAATAATTTTAAAGATTTGTCAGCCCTGCATGATTTATGATGAATCATTTTG-- 587
DB 731 TTTTAAWTTTATWTTWTTWTTWTTTATTTTATTTTATTTTAAATTTTAAAT 672
QY 588 TGGTCTGTAGTTTACTTTTAGAGAGATAGAAAGCATTTGAGGCTCAGGGAAGCAACATT 647
DB 671 TAWWAAATWATTTTATTAATAAATTAATTAATAAATTTTATWAAATAATATTTT 612
QY 648 CAGATGAATCCCAATAGAGAAGGTAAATTTATTTGGGCAATGTACATTTTGGCAGCCTAG 707
DB 611 TTTTATATTAATWATTTATWATWTTTAAATAATWTTTAAATTTTAAATTTTAAATTTTAA 552
QY 708 GCTGTGTACATGTGTACACATTTCTGACATGTGTGTATATGAAATCTTGTCTCTTTT 767
DB 551 ATTATWTTTATTTTATWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 492

Db 809 GKXKKKAAKKKAAADRTTKTKDAAAAAANKTKDVGKKKKKTKTKTKTKKKKKKKKKK 869
 Qy 612 ATAGAAAGCAATTTAGGCTCAGGGAAGCAAAACATTCAGAATGAAATCCAAATAGAGAAGG 671
 Db 869 GGGKDDAAAKKKKGTGKKKKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 928
 Qy 672 TAAATTTATTTGGGCATGTACATTTTGGCAGCCTTAGGCTGTGTACATGTGTACACATTTCT 731
 Db 929 GKXKKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 988
 Qy 732 GAACATGTGTATATTGAAATCTTGTCTCTCTTTTATTGTT 774
 Db 989 KKKKKKKTKTKKKDAAAAATKKKKKKKKTKKKKKKKKKKKKKKKKKKKKKKKKKKK 1031

RESULT 9
 CNS016LI
 LOCUS
 DEFINITION
 CNS016LI melanogaster genome survey sequence 17 end of BAC
 BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epitheroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 TITLE
 SUBMITTED (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr)
 COMMENT
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) at http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
 source
 1. .1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACN16D22"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : 77"

ORIGIN
 Query Match 6.9%; Score 53.8; DB 9; Length 1101;
 Best Local Similarity 30.2%; Pred. No. 0.035;
 Matches 114; Conservative 86; Mismatches 177; Indels 0; Gaps 0;

Qy 297 CCATGTGCATCTAGAAATGAAGTATATTTTAAAAAGGATCGTTTTTGCATGTATAAAT 356
 Db 725 CCCTGWRAGGWWAAAAAATTKATDTTTTTTTTTTTTTTTTTTTTTTTTBTBTATTTT 784
 Qy 357 TTTCAACATTAACCTTCAGGGTATTAATCCCTTTTAAAGTCTAGTTTTTCTTAAGCTG 416
 Db 785 AKDTTTTTTTTATATTTTWTDTTATTTTTTTTAAATATATATTTTAAATATATTA 844
 Qy 417 TGCAGTAATAGAGGTATCGTCATTTCATGTGACATAAAAGATGGAAGGGGCTTCATTCAT 476
 Db 845 WAATTTTATATATATATATATATTTTAAATATATATTTTAAATATATTAATTA 904
 Qy 477 GTTAGTGAAGAAATAGGAAAGTAGGTGAAGTATTTTAAATAGATGTTTCTTTTATGAAA 536
 Db 905 WTWAAWTTTAAWAAWAAATTTATTTWTWATTTATDWRTTTTTASTATTTTDTTWTWATA 964

978	ATAGTGTCTTTTAAATTTATATATATGTAATTTATGTTTATTGTTATTTTATATTTTAT	1037
395	GGCTAGTGTCTTTCTTAAAGTCTGTGCGAGTAATAGAGGTATCGTCATTTCAFGTCACATAAAA	454
1038	TTTATAGTGTCTTTTATTTATTTGTTATATATATTTTATTTGTTATATATTTATTTTATTTATAG	1097
455	GATGGAAGGGGCTTCATTCATGTTAGTCATCGAAATAGGAAAGTAGGTAAGTGAATTTT	514
1098	TTATTTTATTTGTTTATTTATTTGGTATTTATTTATTTTGTATTTATTTATTTTAAATTT	1157
515	AATAGATGTTCTCTTTTATGAAATAATTTTTAAAGATTGTCACGCCCTCATGATTTATG	574
1158	TTTATATATATTTTTTATATATATAGTATTTATTTTATATATATTTATATGATTTAT	1217
575	ATGAATCATTTTGTGTCGTGTTAGTTACTTTTTAGAAAT	613
1218	TTTATAGATATTTTATATATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	1256

RESULT 13					
AU088479/c					
LOCUS	AU088479	500 bp	mRNA	linear	EST 27-JAN-2001
DEFINITION	AU088479	Sugano Malaria	cdna library	Plasmodium falciparum	3D7 cdna
		clone XPfn7173	mRNA sequence.		

VERSION	AU088479.1	GI:12390620
KEYWORDS	EST.	
SOURCE	Plasmodium falciparum 3D7	
ORGANISM	Plasmodium falciparum 3D7	
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS	1 (bases 1 to 500)	
TITLE	Watanabe, J., Sasaki, M., Suzuki, Y. and Sugano, S.	
JOURNAL	FLU-Malaria: a database for a full-length enriched cDNA library	
MEDLINE	from human malaria parasite, Plasmodium falciparum	
PUBMED	Nucleic Acids Res. 29 (1), 70-71 (2001)	
COMMENT	20574754	
	11125052	
	Contact: Junichi Watanabe	
	Institute of Medical Science	
	The University of Tokyo, Department of Parasitology	
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan	
	Tel: 81-3-5449-5378	
	Fax: 81-3-5449-5410	

```

FEATURES
source
Location/Qualifiers
1..500
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/isolate="3D7"
/db_xref="taxon:36329"
/clone="XPFn7173"
/dev_stage="erythrocytic stage"
/clone_lib="Succano Malaria cDNA library"

```

	Query Match	6.6%	Score 51.6;	DB 1;	Length 500;
	Best Local Similarity	48.1%;	Pred. No. 0.1;		
	Matches 138;	Conservative 0;	Mismatches 149;	Indels 0;	Gaps 0;
Qy	336	TCGTTTTGGCGATGATATAAAATTTTCAAACATTAACTTTCAGGGTATTAATCTCTTTAAG	395		
Db	420	TCTTTTTTTACAACCTTTATAAACCTAATCAAAAAATGTTGTGTTTAATATATATAATAAATA	361		
Qy	396	GTCTAGTTTTTCTTAAGTCTGTGCAGTAAATAGAGGTATCGTCAATTCATCTGCACATAAAAG	455		
Db	360	TTATAGGNTATATAAAATTTATCGAAAATTAAGTGTATANATATATATAATAATTTATATAG	301		
Qy	456	ATGGAAAGGGGCTTCATTCTGTTAGTGATGGMAATAGGAAAGTAGGTAGGTCAAGTGATTTTA	515		

```
Db 300 GTTTATAAAATATGNNAAATAAGTCTTANATATATATATATATATATATATATATATATAT 241
Qy 516 ATAGATGTTCTTTATGAATAATATTTTAAAGATGTGTCAGCCCTGATGATTTAAGA 575
Db 240 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 181
Qy 576 TGAATCATCTTTGCGTCTGTTAGTTACTTTTATAGAGATAGAAAGCAT 622
Db 180 TTAATTTAATATTTATTCCTATTTTATTTATATATAAATAATGAT 134

RESULT 14
CNS016BY 1201 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN15C08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106552.1 GI:5622489
VERSION AL106552
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimating the human gene number provided by genome-wide analysis
JOURNAL using Tetraodon nigroviridis DNA sequence
NAT. GENET. 25 (2), 235-238 (2000)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES             Location/Qualifiers
     source           1..1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C08"
                     /clone_lib="DrosBAC"
                     /plasmid="pBelobAC11"
                     /note="end : SP6"

ORIGIN
Query Match          6.6%; Score 51.6; DB 9; Length 1201;
Best Local Similarity 29.5%; Pred. No. 0.12; 199; Indels 0; Gaps 0;
Matches 122; Conservative 92; Mismatches 199; Indels 0; Gaps 0;

Qy 362 AACATACTTTCAGGGTTATTAATCCCTTTAAAGTCTAGTTTTTCTTAAAGTCTGTCAG 421
Db 702 AAGATRGCGMAGTGTBTBTCTTGWAAWAAWAAWAAWADARWADATTTTWTATW 761
Qy 422 TAATAGAGGTATCGTCATTCATGTGACATAAAGATGAAGGGCGTTCATTCATGTAG 481
Db 762 WWWAGNTWTTTTTTTTTTTTTTTTTTAAAAAAMWAAWAAAAAAMAAAGRACCTTGR 821
Qy 482 TGATGGAATAGGAAGTAGGTGAAGTGATTTTATAGATGCTTTCTTTATGAATAATT 541
Db 822 TGGRAGRWGGRARDRGRRAAAWTKDWTGGAWTATAGRAAAAKKRGTAATTTG 881
Qy 542 TTTAAAAAGTTGCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTAGTTA 601
Db 882 GAAADGGGAKDAARRAAGRTTKGKGKRGGTTTDTTTTTTWWAAAAAARWTTTTTWT 941
Qy 602 CTTTTAGAGATAGAAAGCATTTAGGCTCAGGGAAGCAACATTCAGATGAATCAACCA 661
Db 942 TKKKKGGRAATTTTKKGGGTTTCTAADGGGRRKTKGGGRARATTTTCKGWTAARRAG 1001
```

```
Qy 662 ATAGAGAAGCTAAATTTATTTGGCATGTACATTTTGGCAGCCTAGCTGTGTACATCTG 721
Db 1002 TTTTTRKDKKAAAATTTTTTGGRRDGAARAAATKKTTTGTGTAGTGGTGKKKRGKDA 1061
Qy 722 TACACATTTCTGAACATGTGTGTATATTGAAATCTTGTCTCTCTTTTATTTATTT 774
Db 1062 AAAGTAGTWGKTKGKTKKGTWKARARRTTTTTTTTTTTTTTTTTTTTTTTTTT 1114

RESULT 15
CNS04D0K 945 bp DNA linear GSS 01-SEP-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 101H21 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL285149.1 GI:8023560
VERSION AL285149
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimating of human gene number provided by genome-wide analysis
JOURNAL using Tetraodon nigroviridis DNA sequence
NAT. GENET. 25 (2), 235-238 (2000)
COMMENT Estimating Tetraodon nigroviridis DNA sequence
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
JOURNAL
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 945)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source           1..945
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : COBG101CD11LP1-end : T7"

ORIGIN
Query Match          6.6%; Score 51.4; DB 9; Length 945;
Best Local Similarity 39.5%; Pred. No. 0.13; 180; Indels 2; Gaps 1;
Matches 145; Conservative 40; Mismatches 180; Indels 2; Gaps 1;

Qy 300 TGTGTCATCTVAGAAATGAAGTTATATATTTTAAAGAGATCGTTTTTGCATGTATAAATTT 359
Db 801 TTTATATTAAATATTTTTTTTATTTATTTATTTATTTATTTATTTATTTATTTTAT 742
Qy 360 CAAACATTAACCTTCAGGGTTATTAATCCCTTTTAAGGCTAGTTTTTCTTAAGTCTGTC 419
```

Db 741 TATATATTTTTTTTNTNTTATWAWTTTWTWTTTTTTTATTTATTTTATTTTATTTTATTTT 682
QY 420 AGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAGGGGCTTCATTCATGTT 479
Db 681 TATTTTTTATATTTTAWWTTTWTAWWWTTWTATWTTTWTATATAAATATATTATTTTT 622
QY 480 AGTGATGGAATAGGAAAGTAGGTGAAGTGATTTTAAATAGATGTTCTTTTATGAAATAA 539
Db 621 ATWTTTATTTTAA--AAATATTTTWTWTTTTTAAAAATTTTTTATATTATTTTAAATAAAAA 564
QY 540 TTTTAAAGATTTGTCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGT 599
Db 563 ATATWAAAATWTTWTAWAATWTTATTTATTTTAAAAAATTTTATTTTATTTTATTT 504
QY 600 TACTTTTAGAATAGAAAGCATTTGAGGCTCAGGGAAGCAAAACATTCAGAATGAAATC 659
Db 503 TWTWTTTAWAATWNTTTTAAWTTTATTAATTATAAAAAAATAAATAAATAAATAAATAA 444
QY 660 CAATAGA 666
Db 443 AAAAAA 437

Search completed: November 9, 2005, 04:17:53
Job time : 4017.41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:52:53 ; Search time 656.05 Seconds
(without alignments)
9232.381 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125

Perfect score: 125
Sequence: 1 cccctgccccctgccctgcc.....gcgggcagggggcccccagc 125

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_ats.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	125	100.0	308	9 AF012873	AF012873 Homo sapi
2	118.6	94.9	135675	9 AC093174	AC093174 Homo sapi
3	118.6	94.9	180963	9 AC091492	AC091492 Homo sapi
4	117	93.6	148828	9 AY157024	AY157024 Homo sapi
5	62.2	49.8	108399	2 AL356736	AL356736 Homo sapi
6	61	48.8	238794	2 BX936377	BX936377 Homo sapi
7	59.6	47.7	171574	2 AC012300	AC012300 Homo sapi
8	59.4	47.5	135119	2 AC01578	AC01578 Homo sapi
9	58.6	46.9	53121	2 AC136331	AC136331 Homo sapi
10	58.6	46.9	68330	2 AC116109	AC116109 Mus muscu
11	58.2	46.6	72645	2 AC112672	AC112672 Mus muscu
12	58.2	46.6	80542	2 AC020683	AC020683 Homo sapi
13	58	46.4	84701	2 AC016122	AC016122 Homo sapi
14	57.8	46.2	1416	6 AX840138	AX840138 Sequence
15	57.8	46.2	1416	6 AX840139	AX840139 Sequence
16	57.8	46.2	234009	2 CR383672	CR383672 Danio rer
17	57.6	46.1	75449	2 AC016188	AC016188 Homo sapi
18	57.6	46.1	79023	2 AC021951	AC021951 Homo sapi
19	57.4	45.9	59727	2 AC100218	AC100218 Mus muscu

c	20	57.4	45.9	212028	2	CR751564	CR751564 Danio rer
	21	57.4	45.9	218236	2	AC117823	AC117823 Mus muscu
	22	57.4	45.9	240965	2	AC126054	AC126054 Mus muscu
	23	57.2	45.8	87440	2	AC021226	AC021226 Homo sapi
	24	57.2	45.8	218311	2	CR354540	CR354540 Danio rer
	25	56.8	45.4	1060	11	BV111230	BV111230 PZA00983
c	26	56.8	45.4	126667	2	AC015839	AC015839 Homo sapi
	27	56.6	45.3	66680	2	AC100483	AC100483 Mus muscu
	28	56.6	45.3	72847	2	AC023563	AC023563 Homo sapi
	29	56.6	45.3	87283	2	AC022856	AC022856 Homo sapi
c	30	56.6	45.3	180652	2	CR762436	CR762436 Danio rer
	31	56.4	45.1	78700	2	AC022685	AC022685 Homo sapi
	32	56.4	45.1	110000	2	AL929091_06	Continuation (7 of
c	33	56.4	45.1	130244	2	AC013518	AC013518 Homo sapi
c	34	56.4	45.1	186731	2	AC073212	AC073212 Homo sapi
	35	56.4	45.1	194034	9	AL592148	AL592148 Human DNA
	36	56.2	45.0	1703	9	HSHBR2A	X74143 H.sapiens H
	37	56.2	45.0	2904	9	HSHBR2	X78202 H.sapiens H
	38	56.2	45.0	82815	2	AC062001	AC062001 Homo sapi
	39	56	44.8	51274	2	AC023454	AC023454 Homo sapi
	40	56	44.8	229518	2	AC069058	AC069058 Homo sapi
c	41	55.8	44.6	3281	2	AC143602	AC143602 Macaca mu
c	42	55.8	44.6	66596	2	AC131276	AC131276 Homo sapi
c	43	55.8	44.6	78220	2	AC023212	AC023212 Homo sapi
	44	55.8	44.6	81767	2	AC021929	AC021929 Homo sapi
	45	55.8	44.6	132745	2	AC134836	AC134836 Mus muscu

ALIGNMENTS

RESULT 1	AF012873	AF012873	Homo sapiens PPARG gene, isoform 1, promoter and 5'UTR sequence.	308 bp	DNA	linear	PRI 03-MAY-1999
LOCUS	AF012873	AF012873	Homo sapiens PPARG gene, isoform 1, promoter and 5'UTR sequence.	308 bp	DNA	linear	PRI 03-MAY-1999
DEFINITION	AF012873	AF012873	Homo sapiens PPARG gene, isoform 1, promoter and 5'UTR sequence.	308 bp	DNA	linear	PRI 03-MAY-1999
ACCESSION	AF012873	AF012873	Homo sapiens PPARG gene, isoform 1, promoter and 5'UTR sequence.	308 bp	DNA	linear	PRI 03-MAY-1999
VERSION	AF012873.1	GI:4731427	Homo sapiens (human)				
KEYWORDS			Homo sapiens (human)				
SOURCE			Homo sapiens (human)				
ORGANISM			Homo sapiens				
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
			1 (bases 1 to 308)				
			Fajas, L., Auboeuf, D., Raspe, E., Schoonjans, K., Lefebvre, A.-M.,				
			Saladin, R., Najib, J., Laville, M., Fruchart, J.-C., Deeb, S.,				
			Puig-Vidal, A., Flier, J., Briggs, M., Vidal, H., Staels, B. and				
			Auwerx, J.				
TITLE			The organization, promoter analysis, and expression of the human				
			PPARG gene				
JOURNAL			J. Biol. Chem. 272 (30), 18779-18789 (1997)				
MEDLINE			97373577				
PUBMED			9228052				
REFERENCE			2 (bases 1 to 308)				
AUTHORS			Fajas, L., Auwerx, J., Saladin, R. and Briggs, M.				
TITLE			Direct Submission				
JOURNAL			Submitted (07-JUL-1997) Cardiovascular Res., Ligand				
			Pharmaceuticals, 9393 Towne Centre Dr., San Diego, CA 92121-3016,				
			USA				
FEATURES			Location/Qualifiers				
source			1..308				
			/organism="Homo sapiens"				
			/mol_type="genomic DNA"				
			/db_xref="taxon:9606"				
gene			1..>308				
			/gene="PPARG"				
			/note="peroxisome proliferative activated receptor, gamma"				
			1..126				
promoter			/gene="PPARG"				
			127..308				
5'UTR			/gene="PPARG"				
			/note="isoform 1"				
ORIGIN							

FEATURES


```

Db          20832 CC 20833

RESULT 8
AC011578/c
LOCUS      AC011578      135119 bp      DNA      linear      HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-12P9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC011578
VERSION     AC011578.4 GI:9123847
KEYWORDS   HTG; HTGS PHASE0.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 135119)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL   Unpublished
TITLE     2 (bases 1 to 135119)
REFERENCE  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Bouckghalter,B.,
            Brown,A., Castle,A., Collinge,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
            Galagan,J., Gardyna,S., Grant,G., Hegos,B., Heaford,A., Klein,J.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehocsky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (07-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:7107923.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIGR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L3313
            Center clone name: 12_P_9
            -----
            * NOTE: This record contains 146 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1
            * 744: contig of 744 bp in length
            * 745
            * 844: gap of 100 bp
            * 845
            * 1615: contig of 771 bp in length
            * 1616
            * 1715: gap of 100 bp
            * 1716
            * 2485: contig of 770 bp in length
            * 2486
            * 2585: gap of 100 bp
            * 2586
            * 3345: contig of 760 bp in length
            * 3346
            * 4188: contig of 743 bp in length
            * 4189
            * 4288: gap of 100 bp
            * 4289
            * 5050: contig of 762 bp in length
            * 5051
            * 5150: gap of 100 bp
            * 5151
            * 5984: contig of 834 bp in length
            *
            * 5985
            * 6084: gap of 100 bp
            * 6085
            * 6837: contig of 753 bp in length
            * 6838
            * 6937: gap of 100 bp
            * 6938
            * 7694: contig of 757 bp in length
            * 7695
            * 7794: gap of 100 bp
            * 7795
            * 8541: contig of 747 bp in length
            * 8542
            * 8641: gap of 100 bp
            * 8642
            * 9403: contig of 762 bp in length
            * 9404
            * 9503: gap of 100 bp
            * 9504
            * 10273: contig of 770 bp in length
            * 10274
            * 10373: gap of 100 bp
            * 10374
            * 11138: contig of 755 bp in length
            * 11139
            * 11238: gap of 100 bp
            * 11239
            * 11991: contig of 753 bp in length
            * 11992
            * 12091: gap of 100 bp
            * 12092
            * 12840: contig of 749 bp in length
            * 12841
            * 12940: gap of 100 bp
            * 12941
            * 13800: contig of 850 bp in length
            * 13801
            * 13900: gap of 100 bp
            * 13901
            * 14661: contig of 761 bp in length
            * 14662
            * 14761: gap of 100 bp
            * 14762
            * 15508: contig of 747 bp in length
            * 15509
            * 15608: gap of 100 bp
            * 15609
            * 16359: contig of 751 bp in length
            * 16360
            * 16459: gap of 100 bp
            * 16460
            * 17225: contig of 766 bp in length
            * 17226
            * 17325: gap of 100 bp
            * 17326
            * 18074: contig of 749 bp in length
            * 18075
            * 18174: gap of 100 bp
            * 18175
            * 18916: contig of 742 bp in length
            * 18917
            * 19016: gap of 100 bp
            * 19017
            * 19785: contig of 769 bp in length
            * 19786
            * 19885: gap of 100 bp
            * 19886
            * 20641: contig of 756 bp in length
            * 20642
            * 20741: gap of 100 bp
            * 20742
            * 21501: contig of 760 bp in length
            * 21502
            * 21601: gap of 100 bp
            * 21602
            * 22367: contig of 766 bp in length
            * 22368
            * 22467: gap of 100 bp
            * 22468
            * 23213: contig of 746 bp in length
            * 23214
            * 23313: gap of 100 bp
            * 23314
            * 24069: contig of 756 bp in length
            * 24070
            * 24169: gap of 100 bp
            * 24170
            * 24922: contig of 753 bp in length
            * 24923
            * 25022: gap of 100 bp
            * 25023
            * 25771: contig of 749 bp in length
            * 25772
            * 25871: gap of 100 bp
            * 25872
            * 2632: contig of 761 bp in length
            * 2633
            * 26732: gap of 100 bp
            * 26733
            * 27495: contig of 763 bp in length
            * 27496
            * 27595: gap of 100 bp
            * 27596
            * 28366: contig of 771 bp in length
            * 28367
            * 28466: gap of 100 bp
            * 28467
            * 29222: contig of 756 bp in length
            * 29223
            * 29322: gap of 100 bp
            * 29323
            * 30092: contig of 770 bp in length
            * 30192: gap of 100 bp
            * 30193
            * 30945: contig of 753 bp in length
            * 30946
            * 31045: gap of 100 bp
            * 31046
            * 31793: contig of 748 bp in length
            * 31794
            * 31893: gap of 100 bp
            * 31894
            * 32649: contig of 756 bp in length
            * 32650
            * 32749: gap of 100 bp
            * 32750
            * 33490: contig of 741 bp in length
            * 33491
            * 33590: gap of 100 bp
            * 33591
            * 34331: contig of 741 bp in length
            * 34332
            * 34431: gap of 100 bp
            * 34432
            * 35176: contig of 745 bp in length
            * 35177
            * 35276: gap of 100 bp
            * 35277
            * 36039: contig of 763 bp in length
            * 36040
            * 36139: gap of 100 bp
            * 36140
            * 36886: contig of 747 bp in length
            * 36887
            * 36986: gap of 100 bp
            *

```



```

RESULT 10
AC116109
LOCUS
DEFINITION Mus musculus clone RP23-467G14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC116109
VERSION AC116109.1 GI:19703220
KEYWORDS HTG; HTGS PHASE0.
SOURCE
ORGANISM Mus musculus (house mouse)

REFERENCE
1 (bases 1 to 68330)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-467G14
Unpublished
2 (bases 1 to 68330)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kelle,C., LaRocque,K., Lamazates,R.,
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., MacDonald,P., Major,O., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Milhova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,N., Schnapack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22810
Center clone name: 467_G_14
-----
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
* 648 747: gap of 100 bp
* 748 1422: contig of 675 bp in length
* 1423 1522: gap of 100 bp
* 1523 2229: contig of 707 bp in length
* 2230 2329: gap of 100 bp
* 2330 3001: contig of 672 bp in length
* 3002 3101: gap of 100 bp
*
* 3102 3783: contig of 682 bp in length
* 3784 4586: gap of 100 bp
* 4587 4686: gap of 100 bp
* 4687 5381: contig of 695 bp in length
* 5382 5481: gap of 100 bp
* 5482 6190: contig of 709 bp in length
* 6191 6290: gap of 100 bp
* 6291 7003: contig of 713 bp in length
* 7004 7103: gap of 100 bp
* 7104 7824: contig of 721 bp in length
* 7825 7924: gap of 100 bp
* 7925 8662: contig of 738 bp in length
* 8663 8762: gap of 100 bp
* 8763 9402: contig of 640 bp in length
* 9403 9502: gap of 100 bp
* 9503 10235: contig of 733 bp in length
* 10236 10335: gap of 100 bp
* 10336 11055: contig of 720 bp in length
* 11056 11155: gap of 100 bp
* 11156 11861: contig of 706 bp in length
* 11862 11961: gap of 100 bp
* 11962 12653: contig of 692 bp in length
* 12654 13445: contig of 692 bp in length
* 13446 13545: gap of 100 bp
* 13546 14235: contig of 690 bp in length
* 14236 14335: gap of 100 bp
* 14336 15046: gap of 100 bp
* 15047 15146: gap of 100 bp
* 15147 15853: contig of 707 bp in length
* 15854 15953: gap of 100 bp
* 15954 16632: contig of 679 bp in length
* 16633 16732: gap of 100 bp
* 16733 17441: contig of 709 bp in length
* 17442 17541: gap of 100 bp
* 17542 18270: contig of 729 bp in length
* 18271 18370: gap of 100 bp
* 18371 19048: contig of 678 bp in length
* 19049 19148: gap of 100 bp
* 19149 19851: contig of 703 bp in length
* 19852 20661: contig of 710 bp in length
* 20662 20761: gap of 100 bp
* 20762 21473: contig of 712 bp in length
* 21474 21573: gap of 100 bp
* 21574 22286: contig of 713 bp in length
* 22287 22386: gap of 100 bp
* 22387 23086: contig of 700 bp in length
* 23087 23186: gap of 100 bp
* 23187 23920: contig of 734 bp in length
* 23921 24020: gap of 100 bp
* 24021 24713: contig of 693 bp in length
* 24714 24813: gap of 100 bp
* 24814 25540: contig of 727 bp in length
* 25541 25640: gap of 100 bp
* 25641 26364: contig of 724 bp in length
* 26365 26464: gap of 100 bp
* 26465 27180: contig of 716 bp in length
* 27181 27280: gap of 100 bp
* 27281 27977: contig of 697 bp in length
* 27978 28077: gap of 100 bp
* 28078 28769: contig of 692 bp in length
* 28770 28869: gap of 100 bp
* 28870 29559: contig of 690 bp in length
* 29560 29659: gap of 100 bp
* 29660 30357: contig of 698 bp in length
* 30358 30457: gap of 100 bp
* 30458 31165: contig of 708 bp in length
* 31166 31265: gap of 100 bp
* 31266 31948: contig of 683 bp in length
* 31949 32048: gap of 100 bp
* 32049 32768: contig of 720 bp in length

```



```

AC020683      AC020683      80542 bp      DNA      linear      HTG 13-JUL-2000
LOCUS          Homo sapiens clone RP11-30L16, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC020683
AC020683      AC020683      2      GI:9140295
VERSION        HTG: HTGS PHASE0
KEYWORDS
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE         1 (bases 1 to 80542)
JOURNAL
REFERENCE      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
AUTHORS       Homo sapiens, Clone RP11-30L16
TITLE         2 (bases 1 to 80542)
JOURNAL
REFERENCE      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
AUTHORS       Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bada, F.,
              Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castie, A.,
              Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
              Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
              Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
              Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
              Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
              Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
              MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
              McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
              Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K.,
              Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
              Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
              Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
              Tirrell, A., Vasiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
              Zimmer, A. and Zody, M.
              Direct Submission
              Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              On Jul 13, 2000 this sequence version replaced gi:6682660.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: L4895
              Center clone name: 30_L_16
              -----
              * NOTE: This record contains 93 individual
              * sequencing reads that have not been assembled into
              * contigs. Runs of N are used to separate the reads
              * and the order in which they appear is completely
              * arbitrary. Low-pass sequence sampling is useful for
              * identifying clones that may be gene-rich and allows
              * overlap relationships among clones to be deduced.
              * However, it should not be assumed that this clone
              * will be sequenced to completion. In the event that
              * the record is updated, the accession number will
              * be preserved.
              *
              * 1      751: contig of 751 bp in length
              * 752      851: gap of 100 bp
              * 852      1603: contig of 752 bp in length
              * 1604      1703: gap of 100 bp
              * 1704      2545: contig of 842 bp in length
              * 2546      2645: gap of 100 bp
              * 2646      3412: contig of 767 bp in length
              * 3413      3512: gap of 100 bp
              * 3513      4285: contig of 773 bp in length
              * 4286      4386: gap of 100 bp
              * 4386      5154: contig of 769 bp in length
              * 5155      5254: gap of 100 bp
              * 5255      6018: contig of 764 bp in length
              * 6019      6118: gap of 100 bp
              *
              * 6119      6884: contig of 766 bp in length
              * 6885      6984: gap of 100 bp
              * 6985      7755: contig of 771 bp in length
              * 7756      7855: gap of 100 bp
              * 7856      8625: contig of 770 bp in length
              * 8626      8725: gap of 100 bp
              * 8726      9469: contig of 744 bp in length
              * 9470      9569: gap of 100 bp
              * 9570      10438: contig of 769 bp in length
              * 10439      11225: contig of 787 bp in length
              * 11226      11325: gap of 100 bp
              * 11326      12098: contig of 773 bp in length
              * 12099      12198: gap of 100 bp
              * 12199      12973: contig of 775 bp in length
              * 12974      13073: gap of 100 bp
              * 13074      13831: contig of 758 bp in length
              * 13832      13931: gap of 100 bp
              * 13932      14698: contig of 767 bp in length
              * 14699      14798: gap of 100 bp
              * 14799      15569: contig of 771 bp in length
              * 15570      15669: gap of 100 bp
              * 15670      16447: contig of 778 bp in length
              * 16448      16547: gap of 100 bp
              * 16548      17303: contig of 756 bp in length
              * 17304      17403: gap of 100 bp
              * 17404      18172: contig of 769 bp in length
              * 18173      18272: gap of 100 bp
              * 18273      19043: contig of 771 bp in length
              * 19044      19143: gap of 100 bp
              * 19144      19900: contig of 757 bp in length
              * 19901      20000: gap of 100 bp
              * 20001      20769: contig of 769 bp in length
              * 20770      20869: gap of 100 bp
              * 20870      21648: contig of 779 bp in length
              * 21649      21748: gap of 100 bp
              * 21749      22522: contig of 774 bp in length
              * 22523      22622: gap of 100 bp
              * 22623      23389: contig of 767 bp in length
              * 23390      23489: gap of 100 bp
              * 23490      24262: contig of 773 bp in length
              * 24263      24362: gap of 100 bp
              * 24363      25128: contig of 766 bp in length
              * 25129      25228: gap of 100 bp
              * 25229      25989: contig of 761 bp in length
              * 25990      26089: gap of 100 bp
              * 26090      26861: contig of 772 bp in length
              * 26862      26961: gap of 100 bp
              * 26962      27737: contig of 776 bp in length
              * 27738      27837: gap of 100 bp
              * 27838      28612: contig of 775 bp in length
              * 28613      28712: gap of 100 bp
              * 28713      29477: contig of 765 bp in length
              * 29478      29577: gap of 100 bp
              * 29578      30327: contig of 750 bp in length
              * 30328      30427: gap of 100 bp
              * 30428      31195: contig of 768 bp in length
              * 31196      31295: gap of 100 bp
              * 31296      32066: contig of 771 bp in length
              * 32067      32166: gap of 100 bp
              * 32167      32942: contig of 776 bp in length
              * 32943      33042: gap of 100 bp
              * 33043      33818: contig of 776 bp in length
              * 33819      33918: gap of 100 bp
              * 33919      34680: contig of 762 bp in length
              * 34681      34780: gap of 100 bp
              * 34781      35554: contig of 774 bp in length
              * 35555      35654: gap of 100 bp
              * 35655      36412: contig of 758 bp in length
              * 36413      36512: gap of 100 bp
              * 36513      37292: contig of 780 bp in length
              * 37293      37392: gap of 100 bp
              * 37393      38153: contig of 761 bp in length

```


